Predicting difficulty in L2 speech learning

Ocke-Schwen Bohn

Department of English
Aarhus University, Denmark

… by assessing cross-language similarity of speech sounds

This lecture series

Presents an overview of research on
the characteristics, consequences, and causes
of foreign accented speech in perception and in production,
and the methods used to study them.

Stucture:

1. Social, psychological, and communicative consequences of foreign accentedness
2. **Predicting difficulty in L2 speech learning**
3. Core aspects of the revised Speech Learning Model (SLM-r)
4. Do cross-language phonetic relationships provide a full account of L2 speech learning problems?

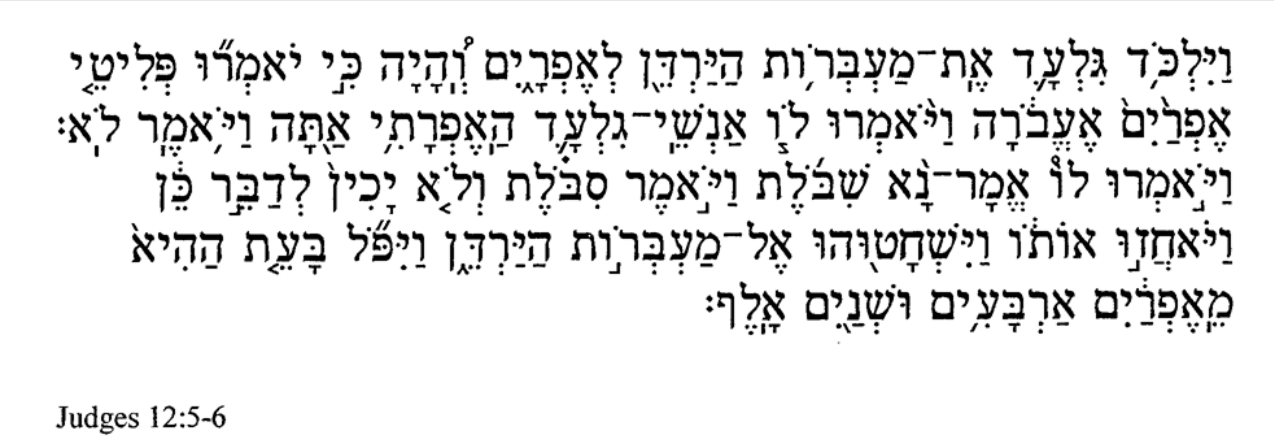
Structure of talk:

1. Difficulty in non-native speech perception and production
2. Approaches to predict difficulty
    - 2.1 Comparison of sound inventories of the L1 and the L2 (based on phonetic/phonemic transcriptions)
    - 2.2 Acoustic comparisons
    - 2.3 Perceptual comparisons
        - 2.3.1 Ecphoric similarity
        - 2.3.2 Perceptual similarity – within and across
3. Conclusion: How to predict difficulty in non-native speech perception and production

Difficulty in non-native speech perception and production

Which difficulty?

Production



וַיִּלְכֹּד גִּלְעָד אֶת־מַעְבְּרוֹת הַיַּרְדֵּן לְאֶפְרָיִם וְהָיָה כִּי יֹאמְרוּ פְלִיטֵי
אֶפְרַיִם אֶעֱבֹרָה וַיֹּאמְרוּ לוֹ אַנְשֵׁי־גִלְעָד הַאֶפְרָתִי אַתָּה וַיֹּאמֶר לֹא׃
וַיֹּאמְרוּ לוֹ אֶמָּר־נָא שִׁבֹּלֶת וַיֹּאמֶר סִבֹּלֶת וְלֹא יָכִין לְדַבֵּר כֵּן
וַיֹּאחֲזוּ אוֹתוֹ וַיִּשְׁחָטוּהוּ אֶל־מַעְבְּרוֹת הַיַּרְדֵּן וַיִּפֹּל בָּעֵת הַהִיא
מֵאֶפְרַיִם אַרְבָּעִים וּשְׁנַיִם אָלֶף׃

Judges 12:5-6

Judges 12

5 The Gileadites captured the fords of the Jordan leading to Ephraim, and whenever a survivor of Ephraim said, "Let me cross over," the men of Gilead asked him, "Are you an Ephraimite?" If he replied, "No," 6 they said, "All right, say 'Shibboleth.' " If he said, "Sibboleth," because he could not pronounce the word correctly, they seized him and killed him at the fords of the Jordan. Forty-two thousand Ephraimites were killed at that time.

1. Difficulty in non-native speech perception and production

Which difficulty?

Perception

Identification of English initial consonants by L1 English listeners

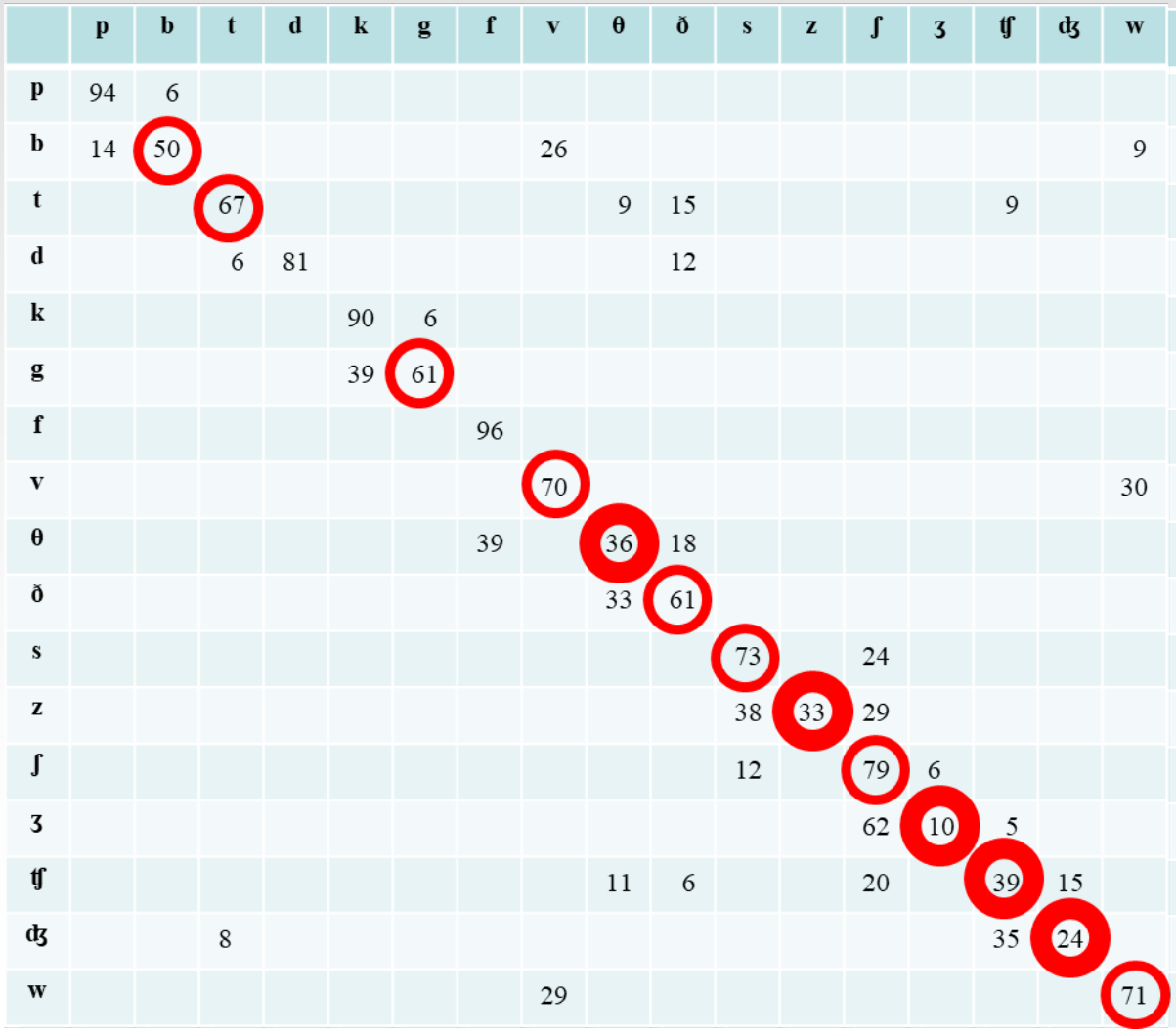
	p	b	t	d	k	g	f	v	θ	ð	s	z	ʃ	ʒ	ʧ	ʤ	w
p	99																
b		99															
t			99														
d				99													
k					99												
g						99											
f							95										
v								91		7							
θ									98								
ð								8		89							
s											99						
z												96					
ʃ													99				
ʒ														88		11	
ʧ															98		
ʤ																88	
w																	99

Difficulty in non-native speech perception and production

Which difficulty?

Perception

Identification of English initial consonants by L1 Kalaallisut (West Greenlandic) listeners



	p	b	t	d	k	g	f	v	θ	ð	s	z	ʃ	ʒ	ʧ	ʤ	w
p	94	6															
b	14	50						26									9
t			67						9	15					9		
d			6	81						12							
k					90	6											
g					39	61											
f							96										
v								70									30
θ							39		36	18							
ð									33	61							
s											73		24				
z											38	33	29				
ʃ											12		79	6			
ʒ													62	10	5		
ʧ									11	6			20		39	15	
ʤ			8												35	24	
w								29									71

Difficulty in non-native speech perception and production

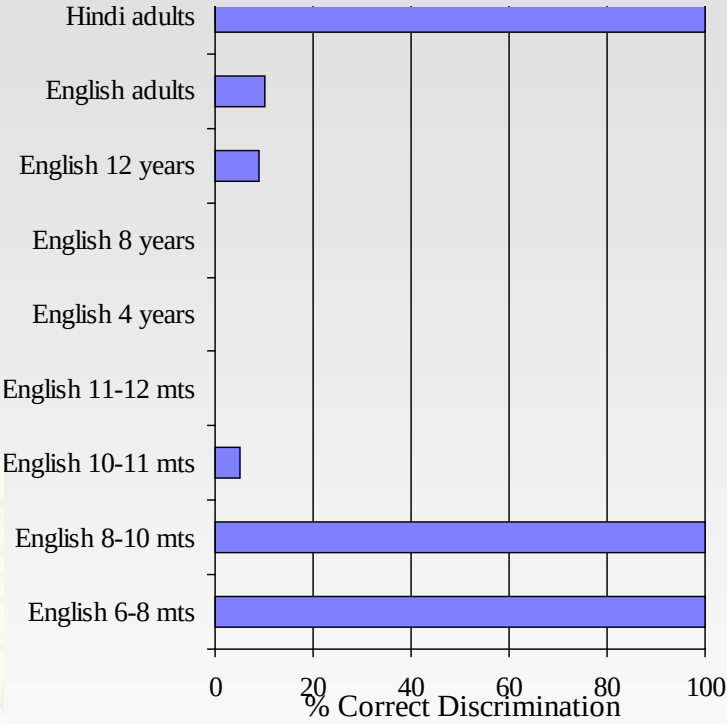
Conclusion: It can be difficult to perceive and to produce nonnative speech sounds correctly

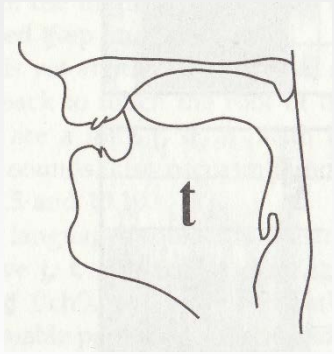
General agreements in L2 speech research:

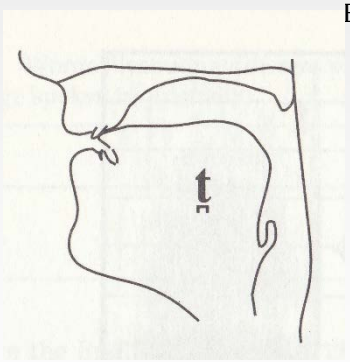
- Origin of difficulty: Rarely (if ever) articulatory
- Cross-language similarity predicts difficulty

Origin of difficulty: Rarely (if ever) articulatory

Example: Perception of Hindi [t̪] - [ʈ]
by L1 English speakers







But: English has [t̪] and [ʈ]
[t̪] in *trunk*, [ʈ] in *eighth*

Origin of difficulty: Rarely (if ever) articulatory

… and certainly unrelated to anatomy:

Retrieved from Seoul Times, 30.9.2012

Cut Tongue for English?

A scene from a movie "If You Were Me" shows that a small child gets ready to undergo a surgery in a Seoul clinic for cutting underneath part of his tongue to make it longer for better pronunciation of English. The surgery called a frenectomy — a minor surgery which lengthens the tongue by about one millimeter, is gaining its popularity among the South Koreans.

Cross-language similarity predicts difficulty

A note on similarity (in general):

”There is nothing more basic to thought and language than our sense of similarity; our sorting of things into kinds"

Quine, W. V. 1969. Natural kinds. In Quine, W. V., *Ontological relativity and other essays*. New York, NY: Columbia University Press, p. 116

In models of cross-language speech perception and of L2 speech learning:

Accuracy of predictions of difficulty in the perception and prodction of nonnative speech sounds depends critically upon valid measures of cross-language similarity.

Importance of measures of cross-language similarity in in models of cross-language speech perception and of L2 speech learning:

Best's Perceptual Assimilation Model (PAM, 1995 and PAM-L2, 2007)

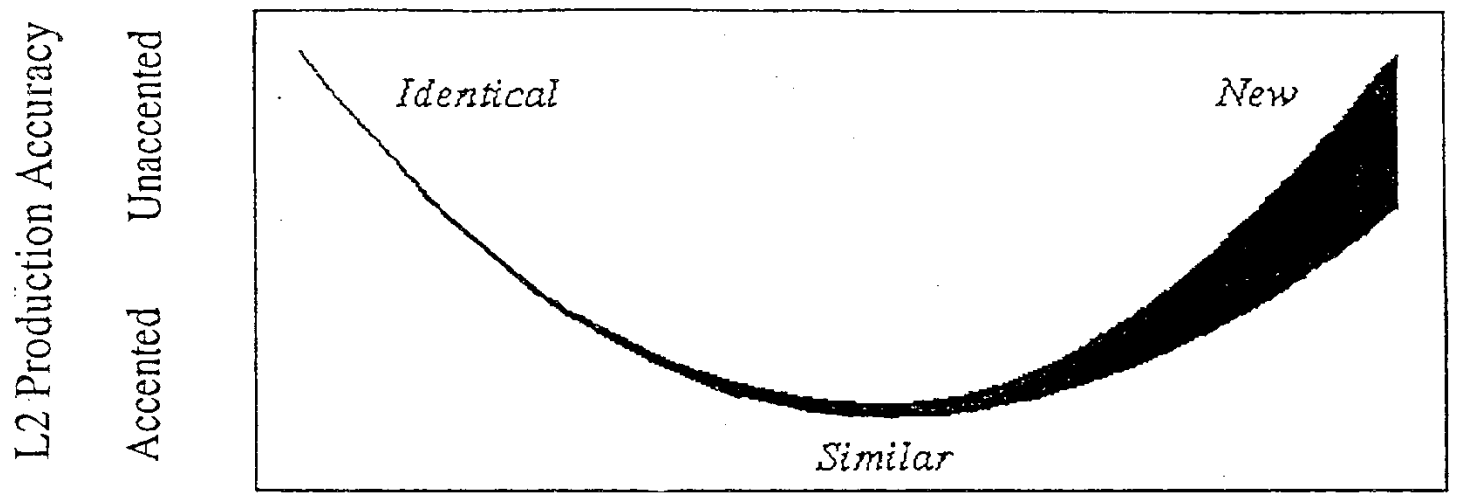
Flege's Speech Learning Model (SLM, 1995 and SLM-r, 2020)

Major's Similarity Differential Rate Hypothesis (SDRH): "The question of how to determine whether something is similar or dissimilar is […] crucial in testing the SDRH" (Major 1997, 221)

Escudero's L2 Linguistic Perception Model (L2LP): "… states that learners will initially perceive L2 sounds in a manner resembling the production of these same sounds in their L1 environment." (van Leussen & Escudero 2015)

The fundamental premise of PAM is that "non-native segments […] tend to be perceived according to their similarities to, and discrepancies from, the native segmental constellations that are in closest proximity to them in the native phonological space" (Best 1995, 193)

Flege's SLM:

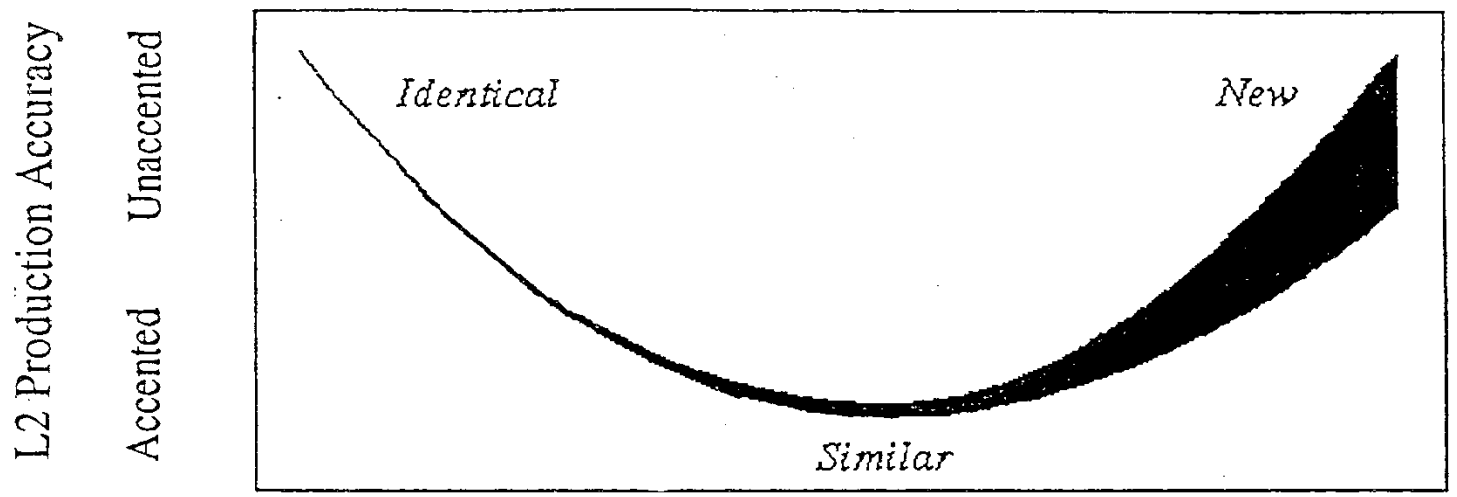


Flege 2005:

Perceived L1 –L2 phonetic dissimilarity regarded as a continuum (since 1994)

(not a tripartite identical-similar-new division, as as from 1984-1993)

Flege's SLM:



Flege 2005:

Perceived L1 –L2 phonetic dissimilarity regarded as a continuum (since 1994)

(not a tripartite identical-similar-new division, as as from 1984-1993)

Must be measured empirically; cannot be predicted *a priori*

⇒ We need valid measures of cross-language similarity of speech sounds

Strange 2007: ”If cross-language phonetic similarity is not empirically characterized *independently* of L2 perceptual difficulties, then the concept becomes circular, with no predictive or explanatory value.”

⇒ We need valid measures of cross-language similarity of speech sounds

Strange 2007: ”If cross-language phonetic similarity is not empirically characterized *independently* of L2 perceptual difficulties, then the concept becomes circular, with no predictive or explanatory value.”

Flege 2005:

What is needed most?

1. More adequate methods of participant selection
2. Standardized measure of perceived L1-L2 phonetic distance

⇒ We need valid measures of cross-language similarity of speech sounds

What about these measures?

Flege (1991, 704): ”**No satisfactory method** now exists for determining whether an L2 vowel will be treated as new or similar.”

Klatt (1987, 781): ”Even the simplest of objectives, such as being able to … relate pairs of vowel spectra in terms of phonetic similarity … are **well beyond our capabilities and understanding**.”

Ladefoged (1987, 3) on ”the problem of how to know when two sounds in different languages should be considered ‘very similar shades of sound' (IPA Principle 2). **I do not know** of any way in which such decisions can be made on theoretical grounds.”

Ladefoged (1990, 344): ”**It is not even technically possible** to devise a measure of auditory distinctiveness among speech sounds without becoming entangled in the problem of observer bias”.

Kohler (1991, 104): ”It is indeed possible to make **language-independent statements** about the auditory distances or similarities of sounds, provided the techniques of investigation … **include the native speaker reaction** in articulation score type experiments, in direct similarity assessment, … for a variety of languages."

2. Approaches to predict difficulty

2.1 Comparison of sound inventories of the L1 and the L2 (based on phonetic/phonemic transcriptions)

2.2 Acoustic comparisons

2.3 Perceptual comparisons

2.3.1 Ecphoric similarity

2.3.2 Perceptual similarity – within and across

DK *det* – DK *dæt*

DK *dæt* – EN *dit*

Comparison of sound inventories of the L1 and the L2 (based on phonetic/phonemic transcriptions)

Typically used as very first approximation:

Same symbols: likely to be similar (identical?) sounds

Different symbols: likely to be dissimilar sounds

Choice of comparing phonemes or allophones across languages pre-empts answer to the empirical question:

At which level is similarity perceived/does interlingual identification occur?

Comparison of sound inventories of the L1 and the L2 (based on phonetic/phonemic transcriptions)

Comparsion of initial consonants of Kalaallisut and English (Bohn & Korneliussen 2019)

14 Kalaallisut consonants					
	Labial	Alveolar	Palatal	Velar	Uvular
Stops	[p]	[t]		[k]	[q]
Fricatives	[v]	[s]		[ɣ]	[ʁ]
Nasals	[m]	[n]		[ŋ]	[ɴ]
Liquids		[l] ~ [ɬ]			
Semivowels			[j]		

24 English Consonants

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Plosive	p b			t d			k g	
Affricate					tʃ dʒ			
Nasal	m			n			ŋ	
Fricative		f v	θ ð	s z	ʃ ʒ			h
Approximant				ɹ		j	w	
Lateral Approximant				l				

Comparison of sound inventories of the L1 and the L2 (based on phonetic/phonemic transcriptions)

14 Kalaallisut consonants					
	Labial	Alveolar	Palatal	Velar	Uvular
Stops	[p]	[t]		[k]	[q]
Fricatives	[v]	[s]		[ɣ]	[ʁ]
Nasals	[m]	[n]		[ŋ]	[ɴ]
Liquids		[l] ~ [ɬ]			
Semivowels			[j]		

24 English Consonants

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Plosive	p b			t d			k g	
Affricate					tʃ dʒ			
Nasal	m			n			ŋ	
Fricative		f v	θ ð	s z	ʃ ʒ			h
Approximant				ɹ		j	w	
Lateral Approximant				l				

	p	b	t	d	k	g
p	94	6				
b	14	50				
t			67			
d			6	81		
k					90	6
g					39	61

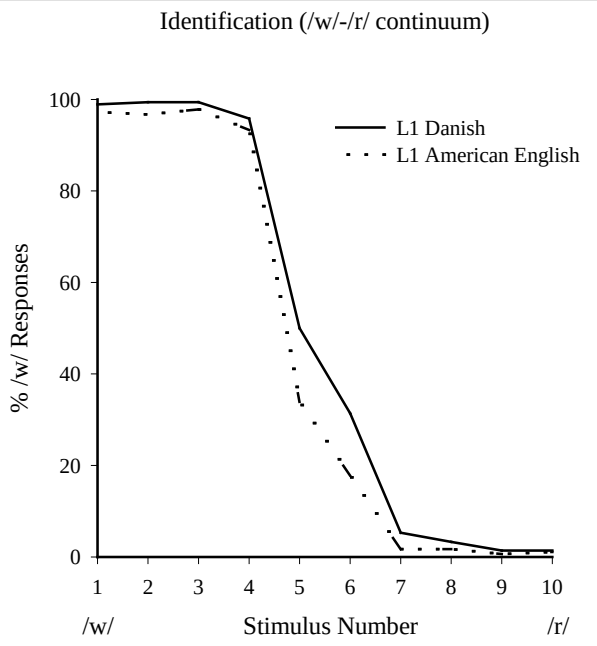
Comparison of sound inventories of the L1 and the L2 (based on phonetic/phonemic transcriptions)

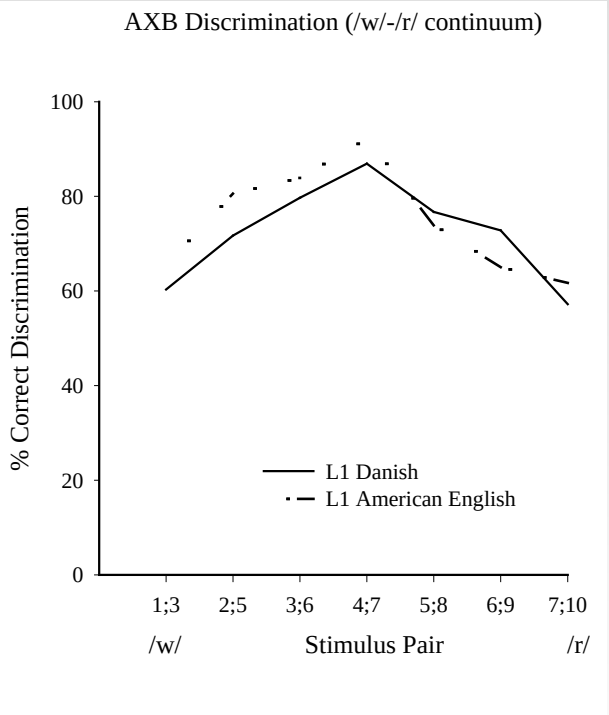
Comparison of approximant phonemes of American English and Danish (Bohn & Best 2012)

AE	DK
/r/	/r/
/l/	/l/
/j/	/j/
/w/	---

Phonologically based prediction: AE /w/-/r/ contrast should be difficult for L1 Danes because they lack /w/

Identification and discrimination of American English /w/-/r/ by L1 Danish listeners





Identification: boundary, slope: n.s.

Discrimination: % correct, % correct (boundary), peakiness: n.s.

Comparison of sound inventories of the L1 and the L2 (based on phonetic/phonemic transcriptions)

Production of Danish /y/
by highly experienced L1 Spanish and L1 English speakers
(Bohn & Garibaldi 2017)

Spanish and English have only two close vowels: /i, u/

Danish has three close vowels: /i, y, u/

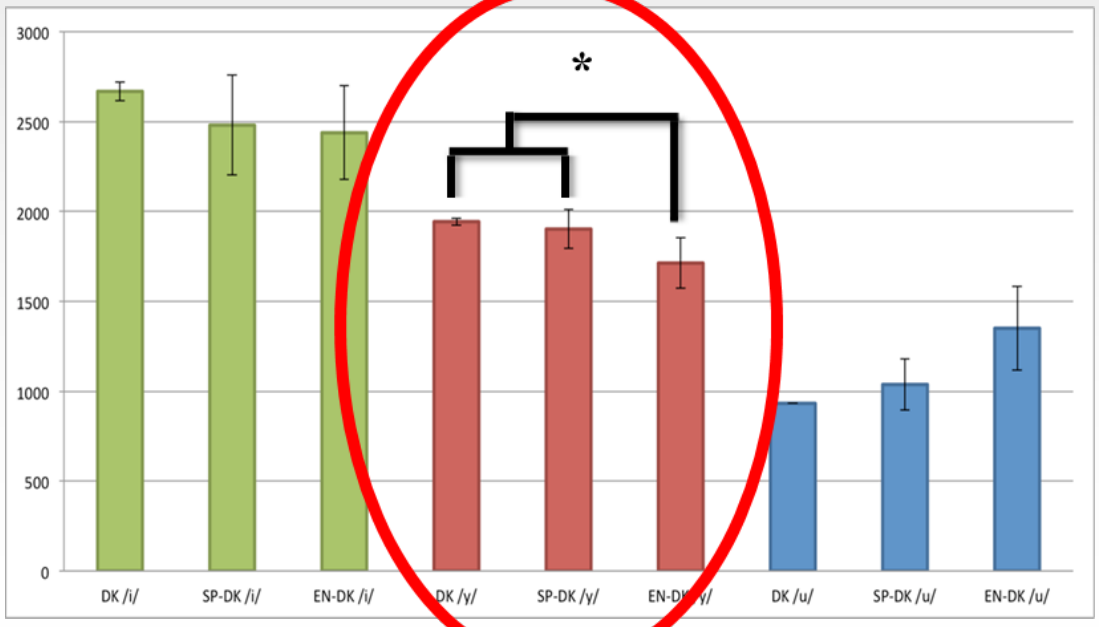
Same problem for L1 speakers of Spanish and of English?

Comparison of sound inventories of the L1 and the L2 (based on phonetic/phonemic transcriptions)

Production of Danish /y/
by L1 speakers of Spanish and of English

L1 with /i, u/ - L2 with /i, y, u/

Main difference between /i, y, u/: Tongue position ~ F2 frequency



Same problem for L1 speakers of Spanish and of English? NO

Comparison of sound inventories of the L1 and the L2 (based on phonetic/phonemic transcriptions)

General problems:

Principles behind use of symbols in the literature are often useless for cross-language comparisons, e.g.,

/hat/ or /hʌt/?

German /a/

Swedish /a/

Southern British English /ʌ/

} identical vowel quality

Comparison of sound inventories of the L1 and the L2 (based on phonetic/phonemic transcriptions)

General problems:

Principles behind use of symbols in the literature are often useless for cross-language comparisons, e.g.,

German /i/-/ɪ/ or /i:/-/ɪ/ or /i:/-/i/ (*biet-bitt*)

compared to

English /i/-/ɪ/ or /i:/-/ɪ/ or /i:/-/i/ (*beat-bit*)

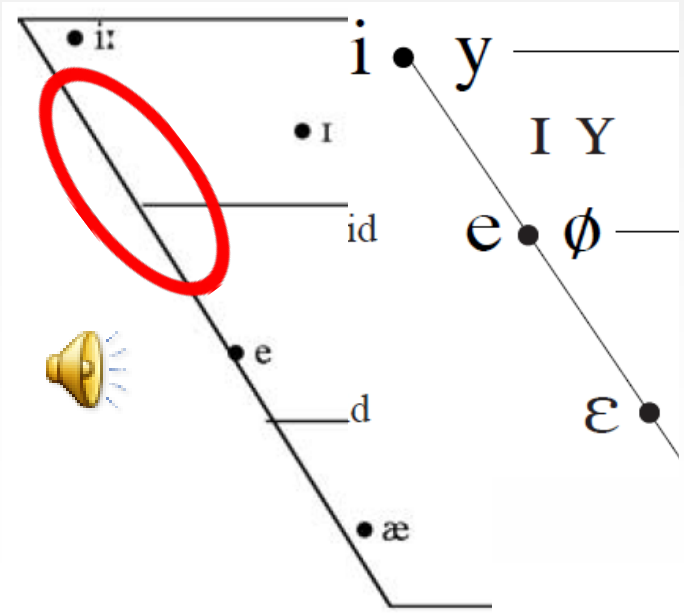
Comparison of sound inventories of the L1 and the L2 (based on phonetic/phonemic transcriptions)

General problems:

Principles behind use of symbols in the literature are often useless for cross-language comparisons, e.g.,

Southern British English /e/ is in fact [ɛ]

head (2 speakers)



Comparison of sound inventories of the L1 and the L2 (based on phonetic/phonemic transcriptions)

General problems:

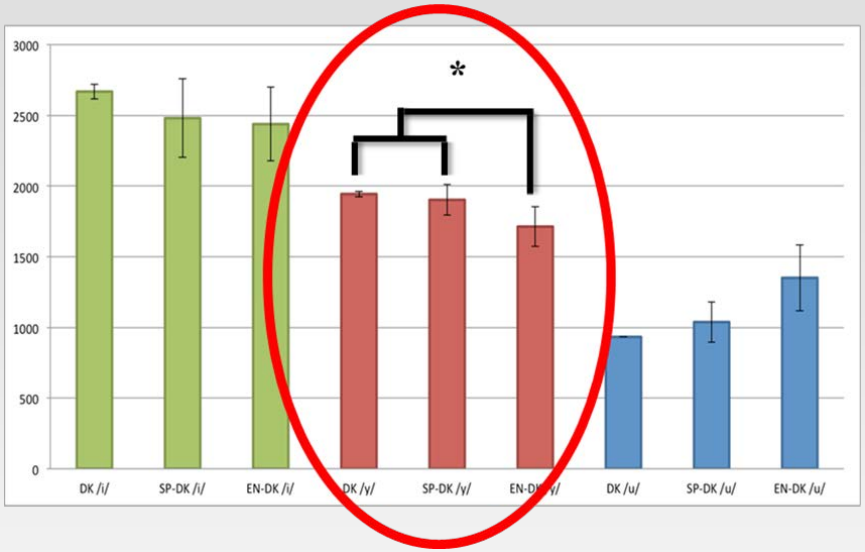
Principles behind use of symbols in the literature are often useless for cross-language comparisons

$\Rightarrow$ Comparison of phonetic symbols can be misleading

Avoid armchair methods inL2 speech research!

2. Approaches to predict difficulty:
Acoustic comparisons

Remember this?



If inventory comparisons do not throw light on why

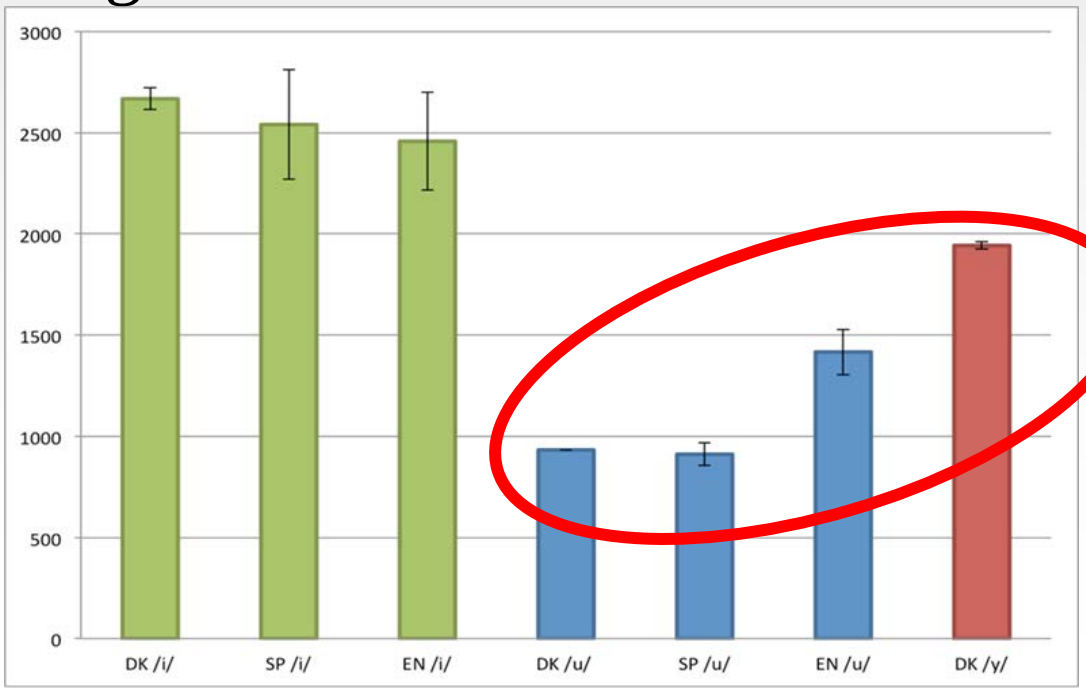
- L1 Spanish speakers produce a Danish-like /y/
- L1 English speakers do not produce a Danish-like /y/

Are acoustic comparisons helpful?

2. Approaches to predict difficulty:
Acoustic comparisons

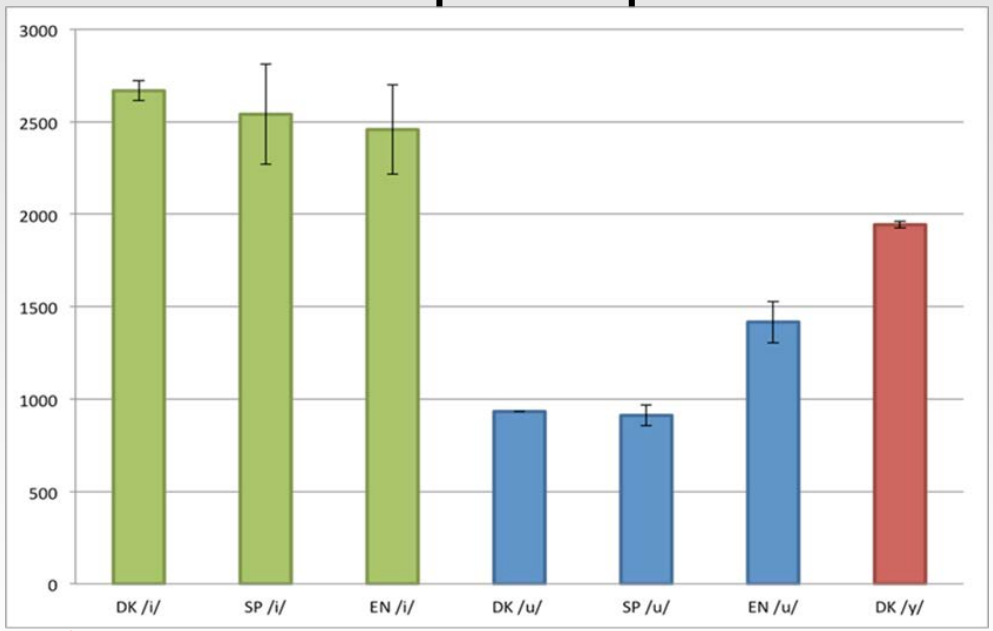
Yes, acoustic comparisons can help predict difficulty in non-native speech production

Acoustic comparison of Danish /y/ to close vowels in Spanish and English



2. Approaches to predict difficulty:
Acoustic comparisons

Yes, acoustic comparisons can help predict difficulty in non-native speech production



Danish /y/ is

- quite dissimilar from any Spanish vowel (in between /i/ and /u/)
- somewhat similar to English /u/

2. Approaches to predict difficulty:
Acoustic comparisons

Yes, acoustic comparisons can help predict difficulty in non-native speech production (one more example later)

BUT: Two major problems with acoustic comparisons

- methodological problems
- problems of validity

2. Approaches to predict difficulty:
Acoustic comparisons

Methodological problems

- Phonetically irrelevant acoustic differences

(due to vocal tract size & shape differences, F0 differences across speaker groups)

-> speaker normalization

(unless point vowels /i a u/ differ only minimally)
(Flege, Munro, Fox 1994)

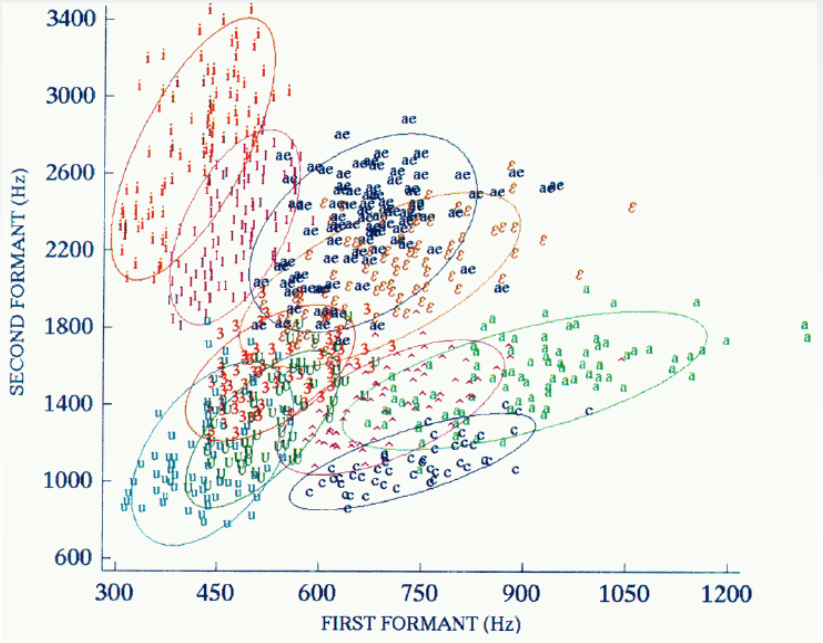


FIG. 4. Values of $F1$ and $F2$ for 46 men, 48 women, and 46 children for 10 vowels with ellipses fit to the data ("æ"=/æ/, "ɑ"=/ɑ/, "ɔ"=/ɔ/, "ʌ"=/ʌ/, "ɝ"=/ɝ/). Measurements for /ɛ/ and /ɔ/ have been omitted, and the data have been thinned of redundant data points.

Hillenbrand et al. 1995: AE vowels produced by 46 men, 48 women, 48 children

2. Approaches to predict difficulty:
Acoustic comparisons

Methodological problems

- Monolingual baseline group = ? L2 target?
- L1 dialectal homogeneity of L2 group?
- context of tokens: “Comparisons of distributions of vowels in multiple phonetic, phonotactic, and prosodic contexts must be performed … “ (Strange 2007)

Acoustic comparisons

Bohn & Flege 1992:

How similar/different are the front vowels of German and English?

Comparison in “identical” contexts, /bVtn/ - /bVt/ in citation form

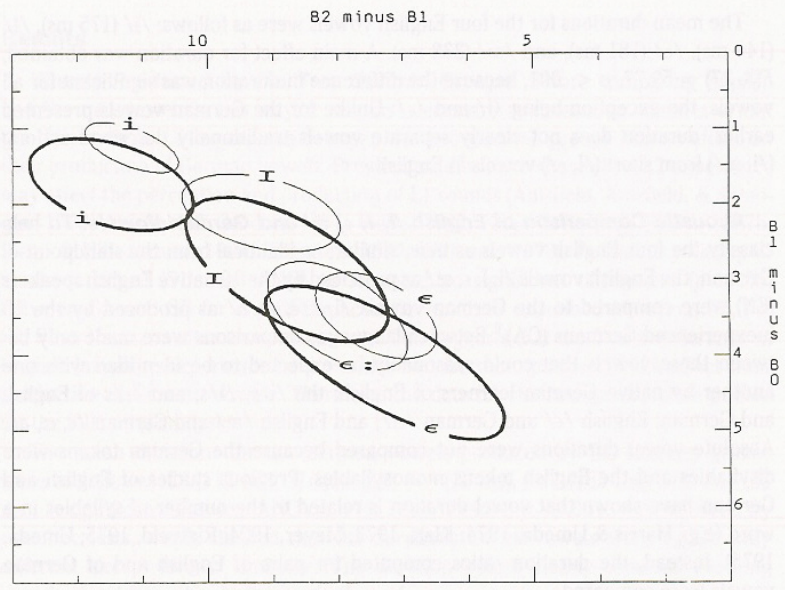


Figure 1. Range of the three English vowels /i, I, ɛ/ (as produced by monolingual native English speakers [bold lines]) and the German vowels /i, I, ɛ, ɛ:/ (as produced by native German speakers with relatively little English language experience [thin lines]) in the Bark-difference space. The ellipses enclose the mean Bark-difference values obtained for the 10 speakers in each group by representing 95% confidence levels based on the two principal components of variation for each vowel.

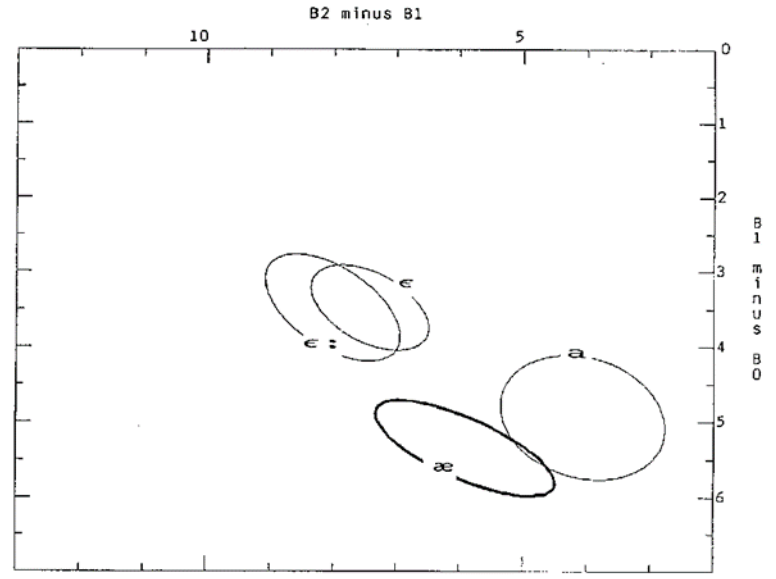


Figure 2. Range of the English vowel /æ/ (as produced by monolingual native English speakers [bold line]) and the German vowels /ɛ, ɛ:, a/ (as produced by native German speakers with relatively little English language experience [thin lines]) in the Bark-difference space.

Acoustic comparisons

North German and American English vowels, /hVt/ (Strange, Bohn, Trent & Nishi 2004):
Result of quantitative assessment of cross-language acoustic similarity (discriminant analysis)

TABLE II. Acoustic similarity (F1/F2/F3 Bark values) of North German (NG) and American English (AE) vowels: Study 1—syllables.

		Modal classification		Other categories	
	NG vowel	AE vowel	# of stimuli	AE vowel	# of stimuli
Front rounded	yː	iː	7	ɪ	1
	øː	ɪ	4	ʊ	2
				uː	2
	ʏ	ɪ	4	ʊ	3
				uː	1
	œ	ɪ	3	ɛ	1
		ʊ	3	ʌ	1

2. Approaches to predict difficulty:
Acoustic comparisons

Summary:

Acoustic comparisons can help predict difficulty in non-native speech production and perception

BUT:

They are methodologically problematic

They can be wrong

Wat nu?

Structure of talk:

1. Difficulty in non-native speech perception and production
2. Approaches to predict difficulty
   - 2.1 Comparison of sound inventories of the L1 and the L2 (based on phonetic/phonemic transcriptions)
   - 2.2 Acoustic comparisons
   - 2.3 Perceptual comparisons
     - 2.3.1 Ecphoric similarity
     - 2.3.2 Perceptual similarity – within and across
3. Conclusion: How to predict difficulty in non-native speech perception and production

Approaches to predict difficulty: Perceived similarity

Perceptual assimilation of the sounds of the L2 to L1 categories ("ecphoric"* similarity)

Discrete labeling tasks (combined with graded ratings of goodness of fit): Subjects label stimuli either

- using an open set of orthographic (or IPA) labels (plus diacritics & verbal comments)

  (e.g., Best, McRoberts & Goodell 2001, Bohn & Best 2012)

- using a closed set of forced choice alternatives

  (e.g., Bohn & Flege 1992; Polka & Bohn 1996; Guion, Flege, Akahana-Yamada & Pruitt 2000; Bohn & Steinlen 2003; Strange, Bohn, Trent & Nishi 2004; Strange, Bohn, Nishi & Trent 2005, Bohn & Best 2012, …)

* Tulving et al. 1983: "Ecphory is a process by which retrieval information provided by a cue is correlated with the information stored in an episodic memory trace …"

Approaches to predict difficulty: Ecphoric similarity

Typical design of perceptual assimilation studies (of ”ecphoric” similarity):

1. Interlingual identification of nonnative stm. with native category
2. Goodness rating of the match (Likert scale)

												passer				
												godt				dårligt
hit	het	hæt	hat	høt	håt	hot	hut	hart	hørt	hårt	(andet)_____	5	4	3	2	1
hit	het	hæt	hat	høt	håt	hot	hut	hart	hørt	hårt	(andet)_____	5	4	3	2	1
hit	het	hæt	hat	høt	håt	hot	hut	hart	hørt	hårt	(andet)_____	5	4	3	2	1

- Ss hear an L2 token
- Ss identify it in terms of representation of L1 category
- Ss rate its goodness of fit to the imagined L1 category

Approaches to predict difficulty: Ecphoric similarity

Perceptual assimilation of Danish [i, y, u] to English /i, u/ and Spanish /i, u/

Result of perceptual assimilation experiment:

1. Mapping of Danish [i, y, u] to native /i/, /u/
2. Rating of goodness of fit of 1. on 1(bad) – 5 (good) Likert scale

Danish stimuli	Spanish response		English response	
	/i/	/u/	/i/	/u/
[i]	100 (3.7)		100 (3.3)	
[y]	33.3 (2.1)	66.7 (2.0)		100 (2.4)
[u]		100 (3.6)		100 (3.2)

Result in terms of **SLM**: Danish [y] is perceived to be more similar to English /u/ by English listeners than to Spanish /u/ by Spanish listeners

SLM prediction: L1 Spanish learners will be ultimately more successful at producing Danish /y/ than L1 English learners

Approaches to predict difficulty: Ecphoric similarity

Perceptual assimilation of Danish [i, y, u] to English /i, u/ and Spanish /i, u/

Result of perceptual assimilation experiment:

1. Mapping of Danish [i, y, u] to native /i/, /u/
2. Rating of goodness of fit of 1. on 1(bad) – 5 (good) Likert scale

Danish stimuli	Spanish response		English response	
	/i/	/u/	/i/	/u/
[i]	100 (3.7)		100 (3.3)	
[y]	33.3 (2.1)	66.7 (2.0)		100 (2.4)
[u]		100 (3.6)		100 (3.2)

Result in terms of **PAM**: Danish [y] is "Uncategorized" for L1 Spanish listeners (does not fit any native category)

PAM prediction: L1 Spanish learners will ultimately be more successful at producing Danish /y/ than L1 English learners

How does ecphoric similarity relate to acoustic similarity?

Perceptual assimilation of North German vowels to American English categories (Strange, Bohn, Trent & Nishi 2004)

TABLE III. Perceptual Assimilation of NG Vowels to AE Categories: Study 1—Syllables.

		Most frequent category			2nd most freque	
	NG vowel	AE vowel	% chosen	Median rating	AE vowel	% chosen
Front rounded	yː	uː	69	2	iː	24
	øː	ʊ	37	1	uː	30
	ʏ	ʊ	56	3	ʌ	20
	œ	ʌ	62	5	ɛ	30

TABLE VII. Perceptual assimilation of NG Vowels to AE Categories: Study 2—Sentence

		Most frequent category			2nd most fre	
	NG vowel	AE vowel	% chosen	Median rating	AE vowel	% cho
Front rounded	yː	uː	87	2	ʊ	7
	øː	uː	43	3	ʊ	2
	ʏ	ʊ	56	2	ʌ	2
	œ	ʌ	80	4	ɛ	1

Compare to: Acoustic similarity of NG and AE vowels

TABLE II. Acoustic similarity (F1/F2/F3 Bark values) of North German (NG) and American English (AE) vowels: Study 1—syllables.

		Modal classification		Other categories	
	NG vowel	AE vowel	# of stimuli	AE vowel	# of stimuli
Front rounded	yː	iː	7	ɪ	1
	øː	ɪ	4	ʊ	2
				uː	2
	ʏ	ɪ	4	ʊ	3
				uː	1
	œ	ɪ	3	ɛ	1
		ʊ	3	ʌ	1

TABLE VI. Acoustic similarity (F1/F2/F3 Bark) of NG and AE vowels: Study 2—Sentences.

		Most frequent		2nd most frequent	
	NG vowel	AE vowel	# of stimuli	AE vowel	# of stimuli
Front rounded	yː	ɪ	6	ɐɪ	2
	øː	ɪ	4	ɐ	2
				ʊ	1
				ɔ	1
	ʏ	uː	4	ʊ	3
				ɪ	1
	œ	ʊ	4	ɛ	2
				ʌ	1
				uː	1

How does ecphoric similarity relate to acoustic similarity?

Acoustic vowel identity is affected by phonetic context
Effects of phonetic contexts are language-specific;
e.g., Bohn 2004:

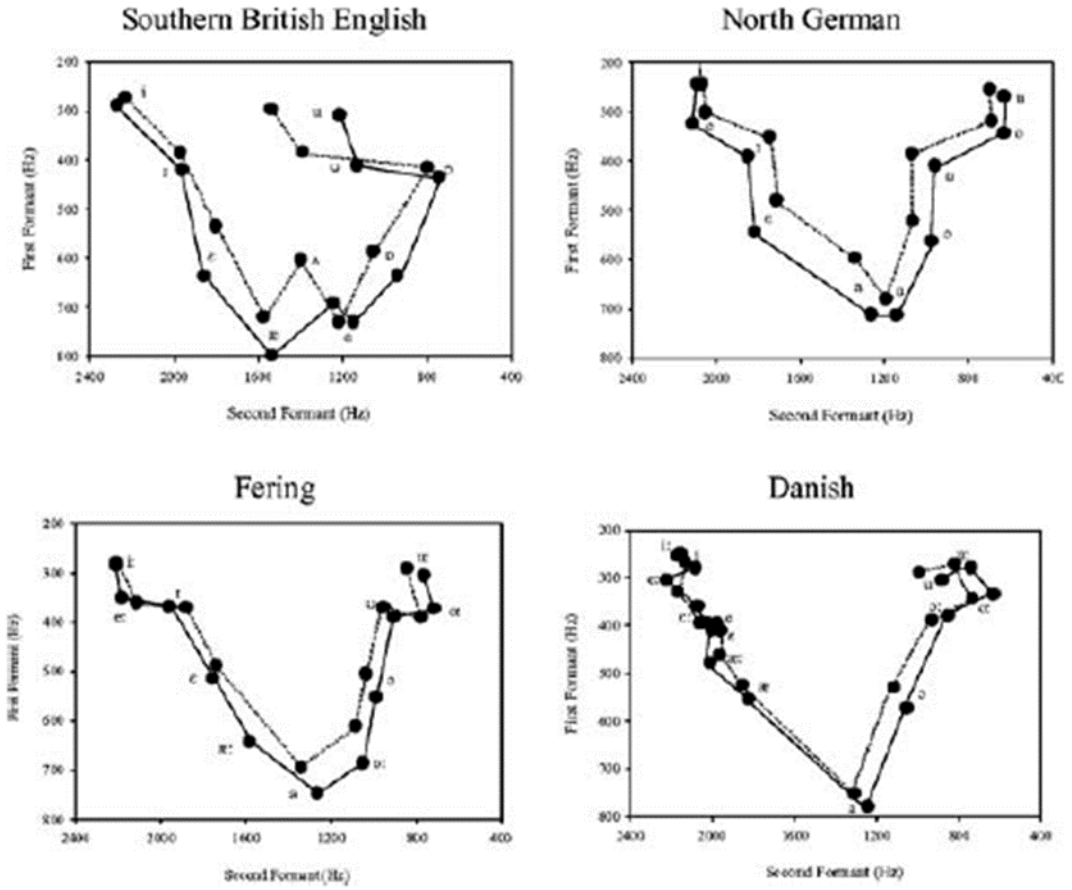


Figure 4 F1/F2 vowel spaces for vowels produced in an /hVt/ (____) and a /dVt/ (. . .) context or four languages differing in vowel inventory size. (Data for English, German, and Danish from Stoeplen 2002)

Do listeners have L1-specific expectations about (degree of) CV-coarticulation which affect cross-language vowel perception?

L1-specific expectations about CV-coarticulation

L1-specific expectations about CV-coarticulation

Strong effects of consonantal context on perceptual assimilation of SBE vowels by Danish listeners (Bohn & Steinlen 2003)

11 SBE monophthongs in /hVt/, /dVt/, /gVk/

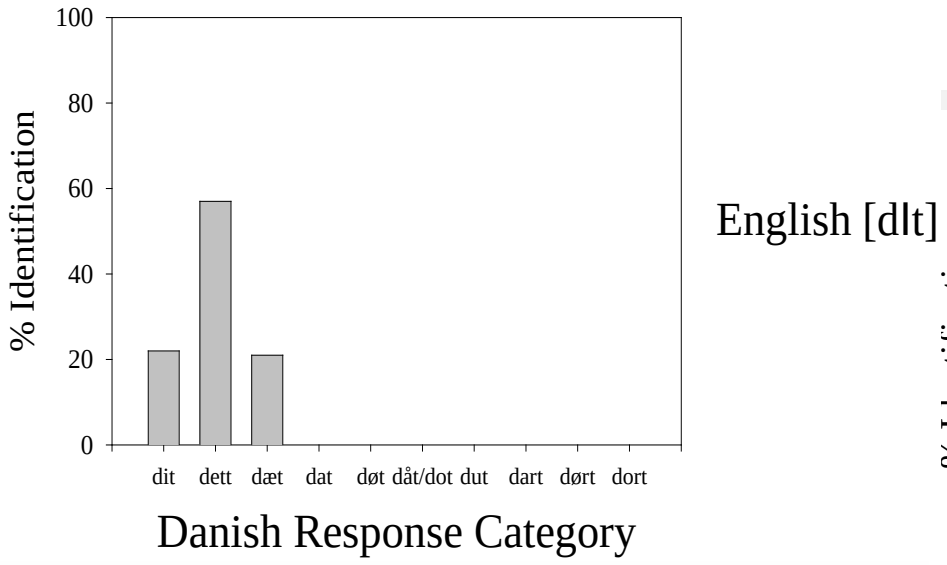
4 talkers x 2 tokens each

10 L1 DK listeners/context

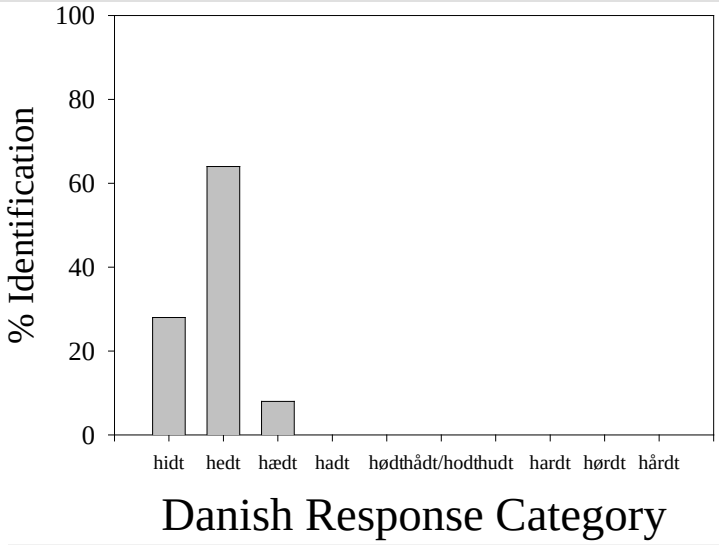
												passer						
												godt						dårligt
hit	het	hæt	hat	høt	hårt	hot	hut	hart	hørt	hårt	(andet)_____	7	6	5	4	3	2	1
hit	het	hæt	hat	høt	hårt	hot	hut	hart	hørt	hårt	(andet)_____	7	6	5	4	3	2	1
hit	het	hæt	hat	høt	hårt	hot	hut	hart	hørt	hårt	(andet)_____	7	6	5	4	3	2	1

L1-specific expectations about CV-coarticulation

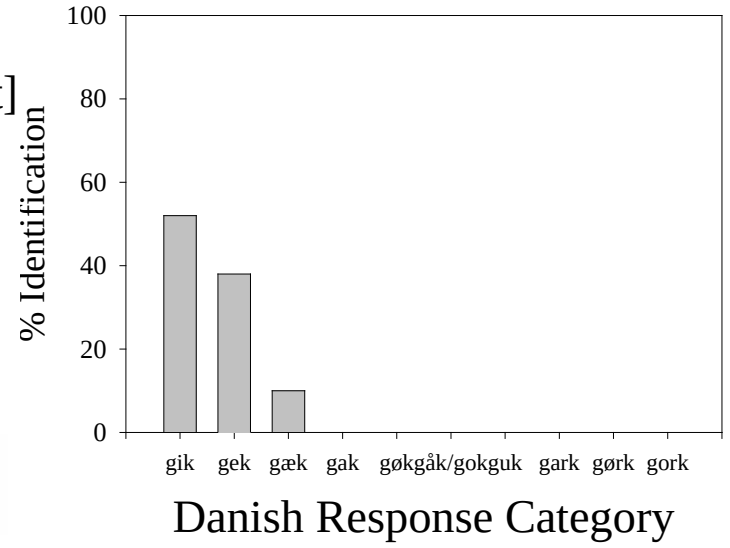
Effects of consonantal context on perceptual assimilation of SBE /ɪ/ by Danish listeners (Bohn & Steinlen 2003)



English [dlt]



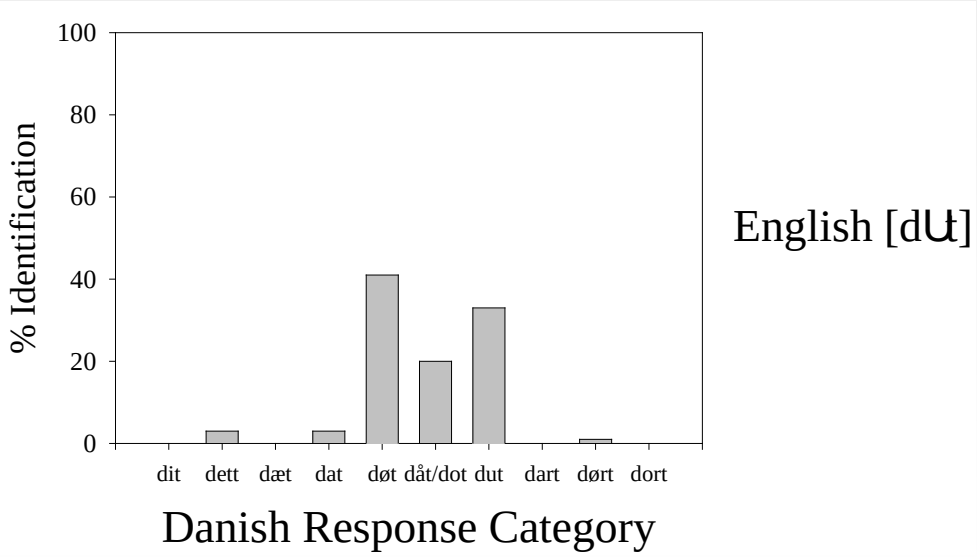
English [hlt]



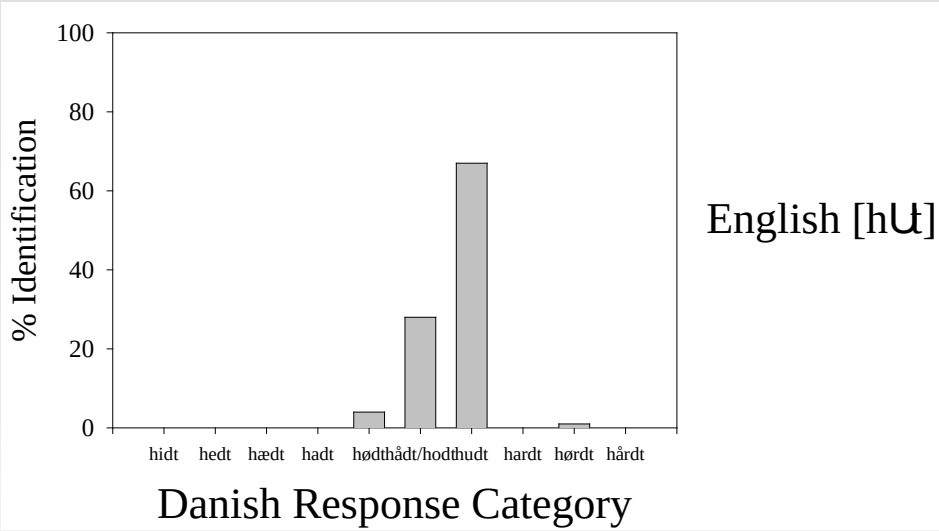
English [glk]

L1-specific expectations about CV-coarticulation

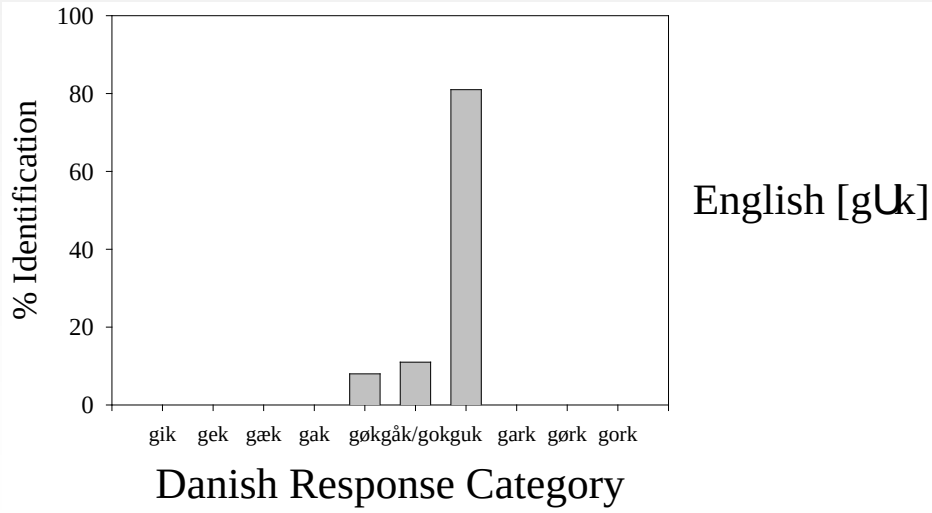
Effects of consonantal context on perceptual assimilation of SBE /ʊ/ by Danish listeners (Bohn & Steinlen 2003)



English [dʊt]



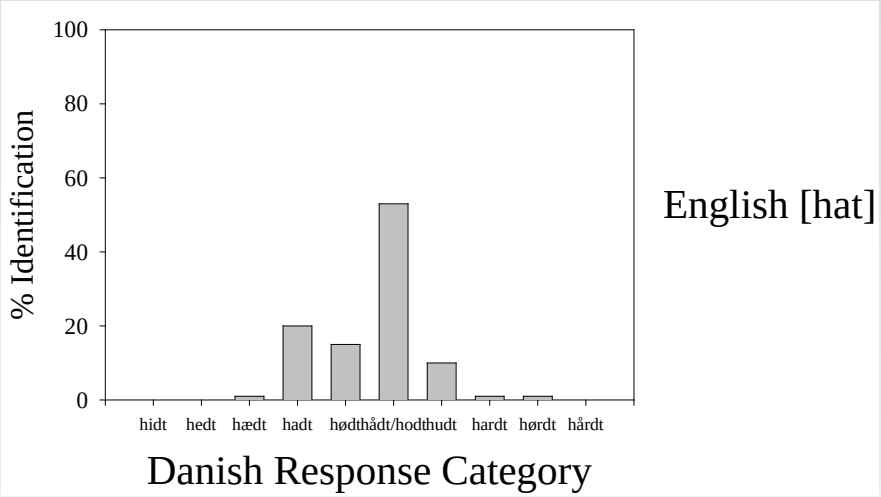
English [hʊt]



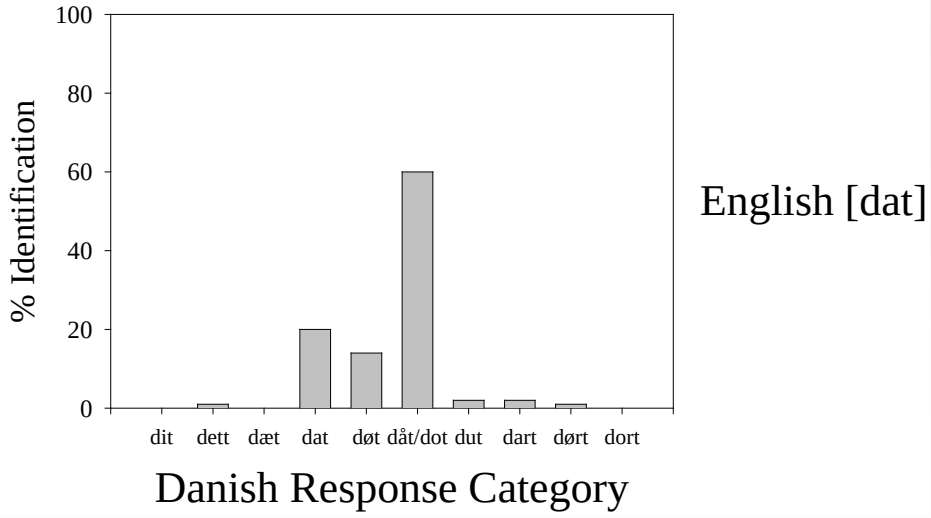
English [gʊk]

L1-specific expectations about CV-coarticulation

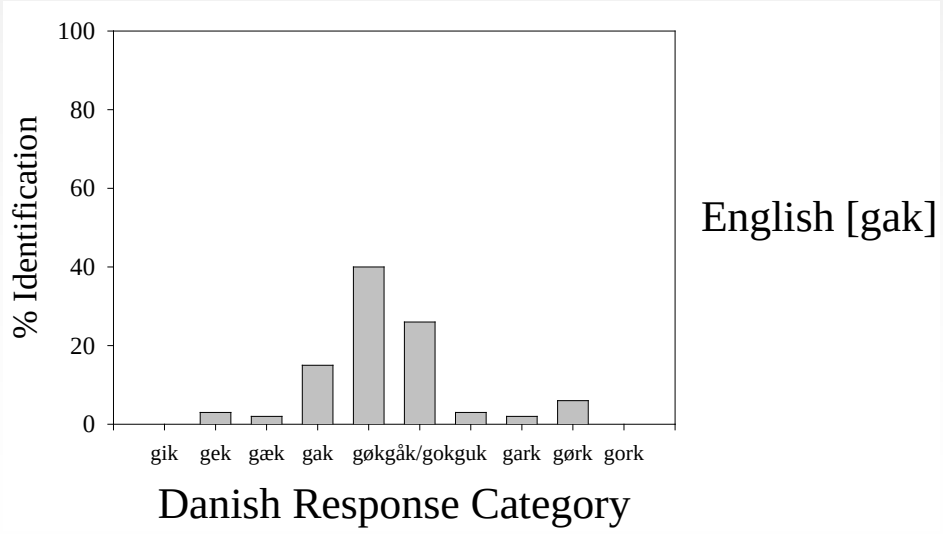
Effects of consonantal context on perceptual assimilation of SBE /ʌ/ by Danish listeners (Bohn & Steinlen 2003)



English [hat]



English [dat]



English [gak]

Approaches to predict difficulty: Ecphoric similarity

So far: Vowels

What about consonants?

Perception and production of English consonants by L1 Danish learners

Production: From Speech Accent Archive http://accent.gmu.edu/

Ca. 3000 samples from 341 L1s of

Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

Approaches to predict difficulty: Ecphoric similarity

Production of English consonants by L1 Danish learners

Speaker Danish 1 from Speech Accent Archive

[plis kɔl stɪlʌ ask hɛ tə bɹɪŋ dis fθ̃ɪŋz̩ fɹʌ̃m̆ wɪd hɛɹ fɹʌ̃m ðə stɔɹ sɪks spũnz əf fɹɛs sno piz̩ faɪf θɪk slæb̥s əv blu tʃiz æ̃n meɪbi ə s̪næk fɔ hɜ bɹʌðɜ b̆ʌp¬ vi ɔlso nid ə smɔl plæstɪk sneɪk æ̃nə bik tɔi fɹɔg fɔ ðə kɪdz̩ si kə̃n skup¬ dis θɪŋz̃ɪntu tɹi ɹɛd bægz̩ æ̃n vi vəl go mit hɛɹ wɛ̃nzdeɪ æt̪ ðə tɹɛ̃ɪn steɪʃə̃n]

Generalizations

Consonant:

- final obstruent devoicing
- interdental fricative to stop

Vowel:

- vowel raising

Approaches to predict difficulty: Ecphoric similarity

Perception of English initial consonants by L1 Danish

Ellegaard & Bohn in prep.

Identified as

inexperienced

Stimuli /Cɑ/

	f	v	θ	ð	s	z	ʃ	ʒ	ʧ	ʤ	w
f	96										
v		57									41
θ	73		22	5							
ð		7	12	72							
s					64	31					
z					35	54	8				
ʃ							63	25	8		
ʒ							26	37		27	
ʧ			6				6		83		
ʤ									25	72	
w		20									80

Approaches to predict difficulty: Ecphoric similarity

Perceptual assimilation English -> Danish initial consonants

inexperienced

English Stimuli

Assimilated to Danish consonants

	f	v	s	ɕ	tj	dj	w	NO FIT
f	100 (7.2)							
v		63 (6.1)					38 (6.0)	
θ	96 (6.4)							
ð		20 (4.3)			10 (3.2)	28 (3.2)		33
s			97 (6.7)					
z			93 (5.7)	6 (4.9)				
ʃ				98 (6.6)				
ʒ				54 (5.1)	7.3 (3.1)	30 (3.4)		
ʧ				9 (6.6)	90 (6.2)			
ʤ					40 (6.8)	58 (7.0)		
w		14 (4.7)					86 (6.4)	

Approaches to predict difficulty: Ecphoric similarity

Does ecphoric similarity predict identification problems?

Zoom in on /f, θ, ð/

Perceptual assimilation
EN -> DK

	f	v	tj	dj	w	NO FIT
f	100 (7.2)					
v		63 (6.1)			38 (6.0)	
θ	96 (6.4)					
ð		20 (4.3)	10 (3.2)	28 (3.2)		33

Identification of
EN /f, θ, ð/
by L1 DK

	f	v	θ	ð	w
f	96				
v		57			41
θ	73		22	5	
ð		7	12	72	

Approaches to predict difficulty: Ecphoric similarity and within-language discrimination

Bohn & Ellegaard 2019: Does graded within-language discrimination predict identification problems?

L1 Danish listeners rated

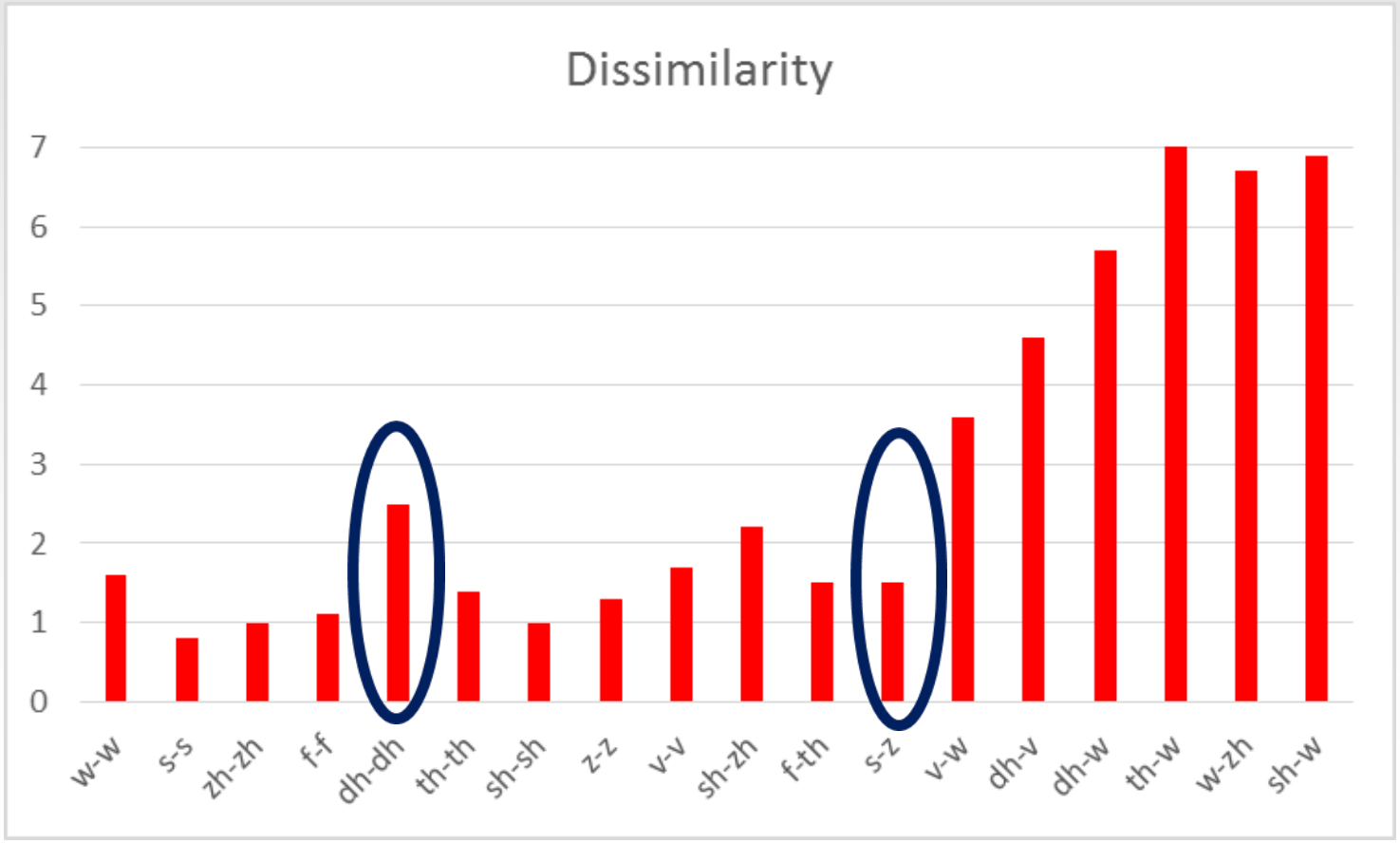
9 EN consonant contrasts [ʃ-ʒ], [f-θ], [s-z], [v-w], [ð-v], [ð-w], [θ-w], [w-ʒ], [ʃ-w]

and 9 foils [w]-[w], [s]-[s], [ʒ]-[ʒ], [f]-[f], [ð]-[ð], [θ]-[θ], [ʃ]-[ʃ], [z]-[z], and [v]-[v]

for (dis-)similarity on 7-point Likert scale

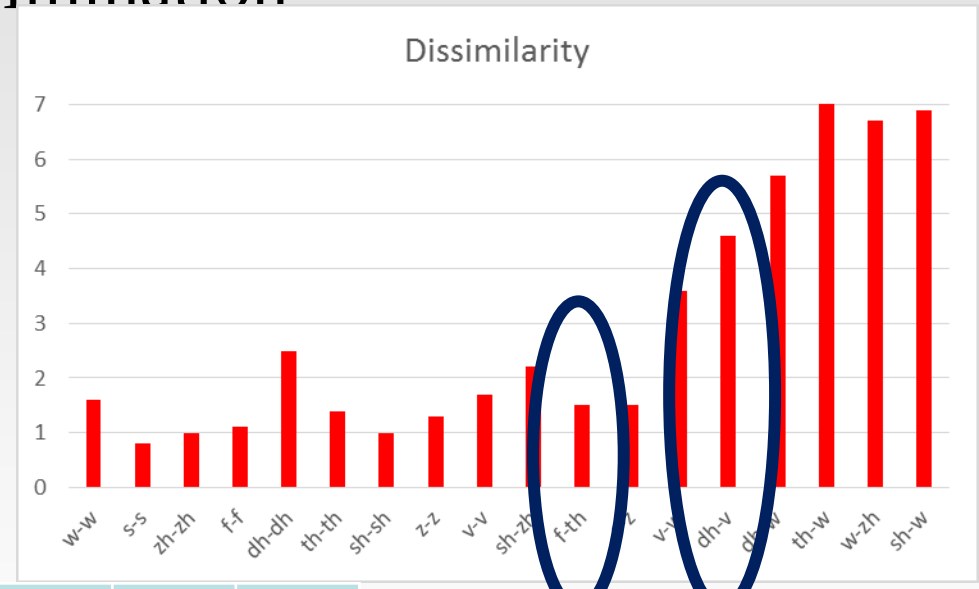
Approaches to predict difficulty: Ecphoric similarity and within-language discrimination

Graded within-language discrimination



Approaches to predict difficulty: Ecphoric similarity and within-language discrimination

Graded within-language discrimination



Zoom in on /f, θ, v, ð/

Identification of EN /f, θ, v, ð/ by L1 DK

	f	v	θ	ð	w
f	96				
v		57			41
θ	73		22	5	
ð		7	12	72	

Approaches to predict difficulty: Ecphoric similarity and within-language discrimination

Conclusion:

Both ecphoric similarity (perceptual assimilation task)

and

within-L2 discrimination

predict learning difficulty quite well

Approaches to predict difficulty: An alternative to ecphoric similarity

However

Ecphoric similarity (perceptual assimilation task) can be problematic to interpret:

Current typical procedure:

												passer						
												godt						dårligt
hit	het	hæt	hat	høt	håt	hot	hut	hart	hørt	hårt	(andet)_____	7	6	5	4	3	2	1
hit	het	hæt	hat	høt	håt	hot	hut	hart	hørt	hårt	(andet)_____	7	6	5	4	3	2	1

- Ss hear an L2 token
- Ss identify it in terms of imagined L1 category
- Ss rate its goodness of fit to the imagined L1 category

What are subjects thinking about?

Questions about the interpretation of ecphoric similarity ratings

What is the nature of the rating scale?

															passer			
												godt						dårligt
hit	het	hæt	hat	høt	hårt	hot	hut	hart	hørt	hårt	(andet)_____	7	6	5	4	3	2	1
hit	het	hæt	hat	høt	hårt	hot	hut	hart	hørt	hårt	(andet)_____	7	6	5	4	3	2	1

Interval? $\Rightarrow$ Ratings can be treated parametrically (central tendency = mean)

Questions about the interpretation of ecphoric similarity ratings

Example of parametrically treated goodness ratings:

Guion, Flege, Akahane-Yamada & Pruitt 2000:

Fit index: Consistency score x goodness rating

TABLE III. Fit indexes derived for English consonants in terms of Japanese categories. The fit index was derived from the proportion of identifications and goodness ratings (see the text). Only identifications that were more than 30% are included.

English consonant	Most common identification	Proportion of identifications	Goodness rating	Fit index	
/b/	/b/	0.84	5.3	4.5	good /b/
/s/	/s/	0.87	4.5	3.9	good /s/
/t/	/t/	0.91	3.9	3.5	good /t/
/v/	/v/	0.80	4.4	3.5	good /v/
/w/	/ɯ/	0.79	3.5	2.8	fair /ɯ/
/ɹ/	/ɯɾ/	0.50	3.3	1.7	poor /ɯɾ/
	/ɾ/	0.46	3.4	1.6	poor /ɾ/
/l/	/ɾ/	0.50	3.2	1.6	poor /ɾ/
	/ɯɾ/	0.37	3.0	1.1	poor /ɯɾ/
/θ/	/s/	0.39	3.8	1.5	poor /s/
	/ɸ/	0.38	3.4	1.3	poor /ɸ/

Questions about the interpretation of ecphoric similarity ratings

What is the nature of the rating scale?

														passer				dårligt
												godt						
hit	het	hæt	hat	høt	hårt	hot	hut	hart	hørt	hårt	(andet)_____	7	6	5	4	3	2	1
hit	het	hæt	hat	høt	hårt	hot	hut	hart	hørt	hårt	(andet)_____	7	6	5	4	3	2	1

Interval? ⇒ Ratings can be treated parametrically
(central tendency = mean)

Ordinal? ⇒ Ratings must be treated nonparametrically
(central tendency = median or mode, as in Strange et al. 2004, 2005)

Approaches to predict difficulty: An alternative to ecphoric similarity

However

Ecphoric similarity (perceptual assimilation task) can be problematic to interpret:

Current typical procedure:

												passer						
												godt						dårligt
hit	het	hæt	hat	høt	håt	hot	hut	hart	hørt	hårt	(andet)_____	7	6	5	4	3	2	1
hit	het	hæt	hat	høt	håt	hot	hut	hart	hørt	hårt	(andet)_____	7	6	5	4	3	2	1

- Ss hear an L2 token
- Ss identify it in terms of imagined L1 category
- Ss rate its goodness of fit to the imagined L1 category

What are subjects thinking about?

Approaches to predict difficulty: An alternative to ecphoric similarity

Well, what are subjects thinking about?

For some nonnative listeners, we will never know:

In the Strange et al. studies, ca. 1/3 of (potential) participants had to be excluded:

Even after 50-100 training trials:
unable to use English orthographic symbols reliably

-> General problem of orthographic representation of speech sounds in some languages

An alternative to ecphoric similarity?

1. Discrete labeling tasks revealing *”ecphoric”* similarity (combined with graded ratings of goodness of fit): Listeners label stimuli

2. Graded discrimination tasks revealing *”perceptual”* similarity:

   Listeners compare similarity of 2 auditorily presented stimuli (e.g., Flege, Munro & Fox 1994)

   Designed not to access listeners’ own internal representations of phonetic categories (→ ”ecphoric” similarity)

   Designed to compare specific productions of L2 and L1 phones

DK

det ***dæt*** ***dat***

SBE

dit ***debt***

Comparison of ecphoric and perceptual similarity

How does ecphoric similarity

(as examined in perceptual assimilation tasks)

relate to

perceptual similarity

(as examined in ”overt” tasks)?

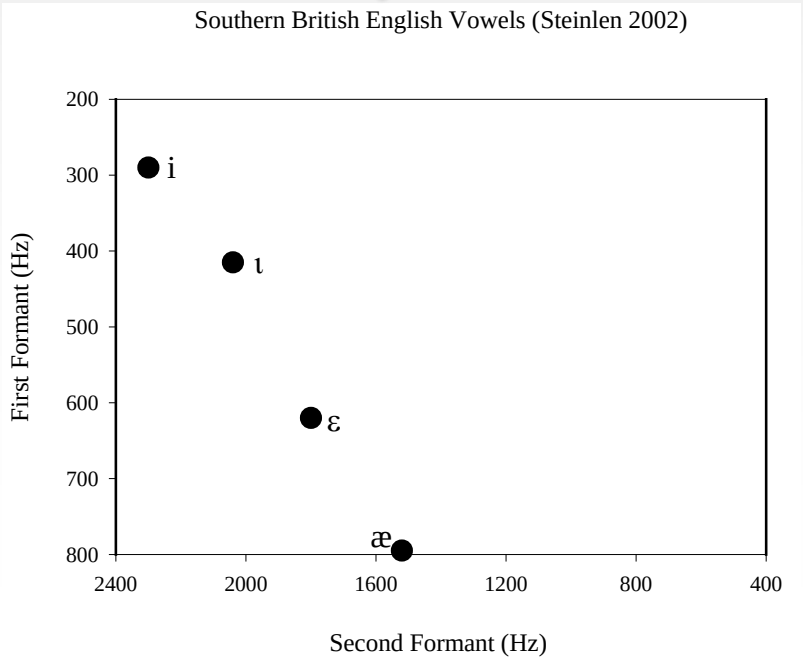
Comparison of ecphoric and perceptual similarity

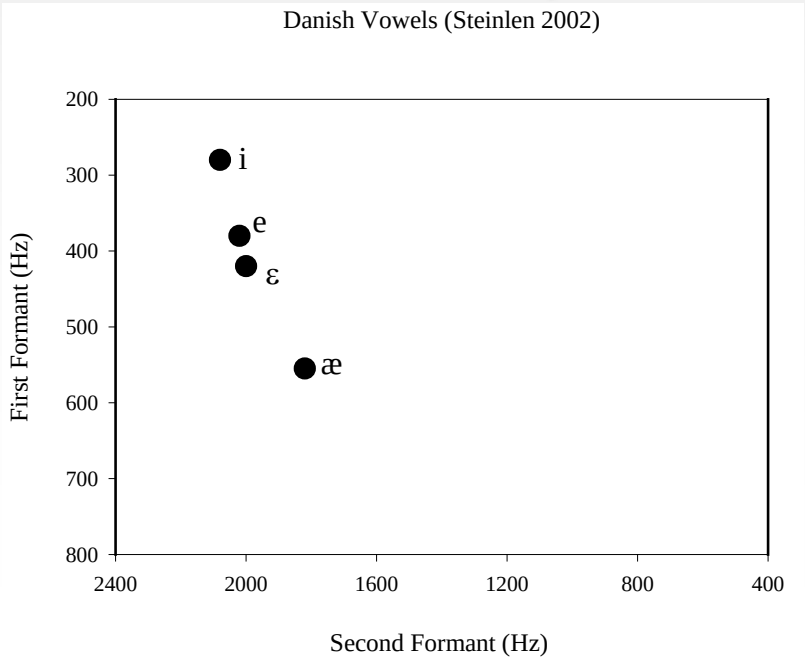
Garibaldi & Bohn in prep.

Comparison of ecphoric and perceptual similarity of Southern British English and Danish front vowels

/i, ɪ, ɛ, æ/

/i, e, ɛ, æ/





Experiment 1: Ecphoric similarity

Subjects: 10 L1 SBE listeners (2f, 8m, M_{age} = 29.1), minimal L2 Danish experience)

Stimuli: 4 (short) front unrounded vowels of Danish (i, e, ɛ, æ) in /hVt/

4 male L1 Danish speakers, 5 repetitions = 80 trials

Tasks:

1. Interlingual identification of nonnative stm. with native category
2. Goodness rating of the match (Likert scale)

				goodness of fit				
				bad				perfect
heat	hit	het	hat	1	2	3	4	5

Experiment 2: Perceptual similarity

Subjects: 10 L1 SBE listeners (2f, 8m, M_{age} = 29.1), minimal L2 Danish experience)

Stimuli: 4 front (unrounded) monophthongs of SBE (i, ɪ, ɛ, æ) and DK (i, e, ɛ, æ) in /hVt/, produced by 4 male L1 speakers each, randomly presented in 16 SBE-DK and 16 DK-SBE pairs plus 8 ”catch” pairs, ISI = 1.0 s, each pair 4 x = 160 trials

SBE vowels →

DK vowels↓

	/i/	/ɪ/	/ɛ/	/æ/
/i/				
/e/				
/ɛ/				
/æ/				

Task:

Graded discrimination using Likert-scale:

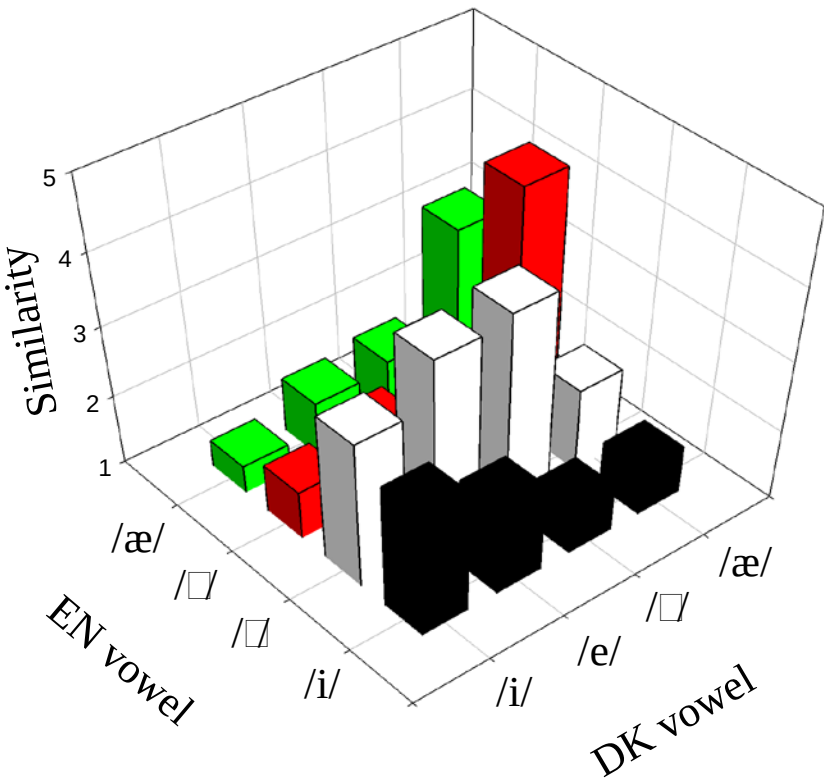
1 (”very different”) ……..5 (”same”)

Comparison of perceptual and ecphoric similarity

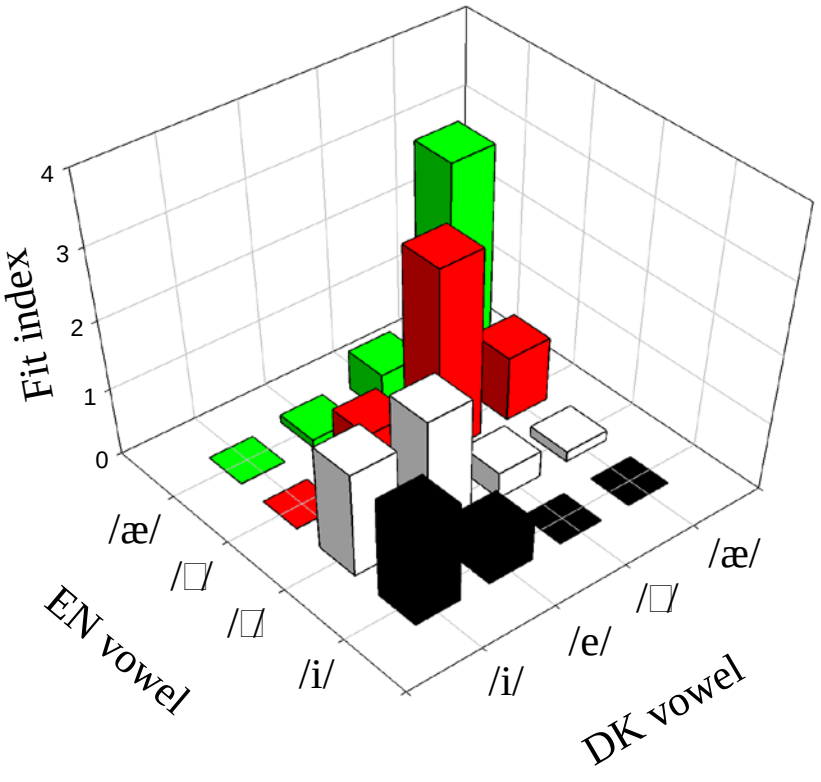
Graded discrimination

Fit index
p identification x goodness rating

Perceptual similarity
Danish front vowels <-> English front vowels
/hVt/



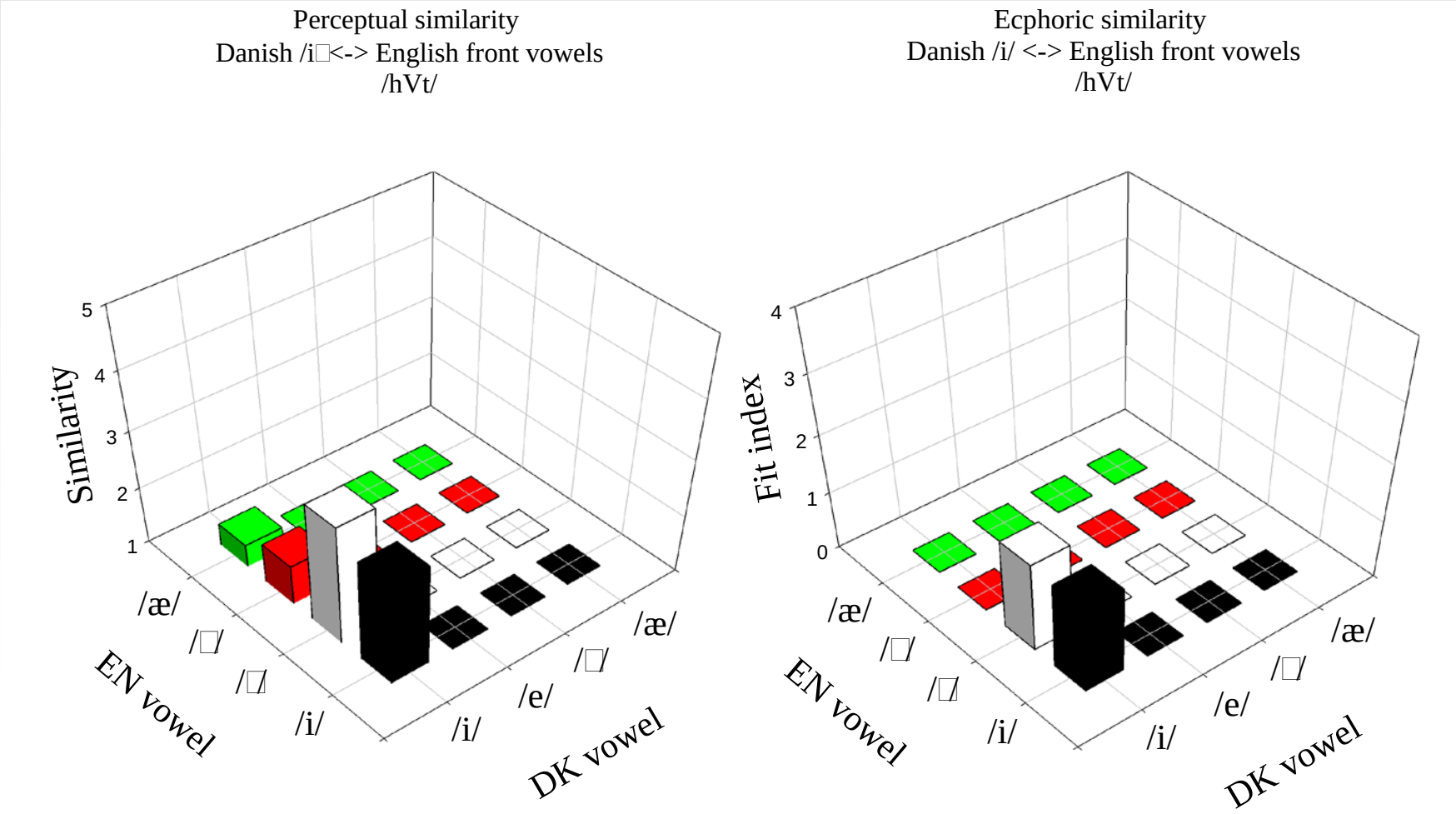
Ecphoric similarity
Danish front vowels <-> English front vowels
/hVt/



Comparison of perceptual and ecphoric similarity

Graded discrimination

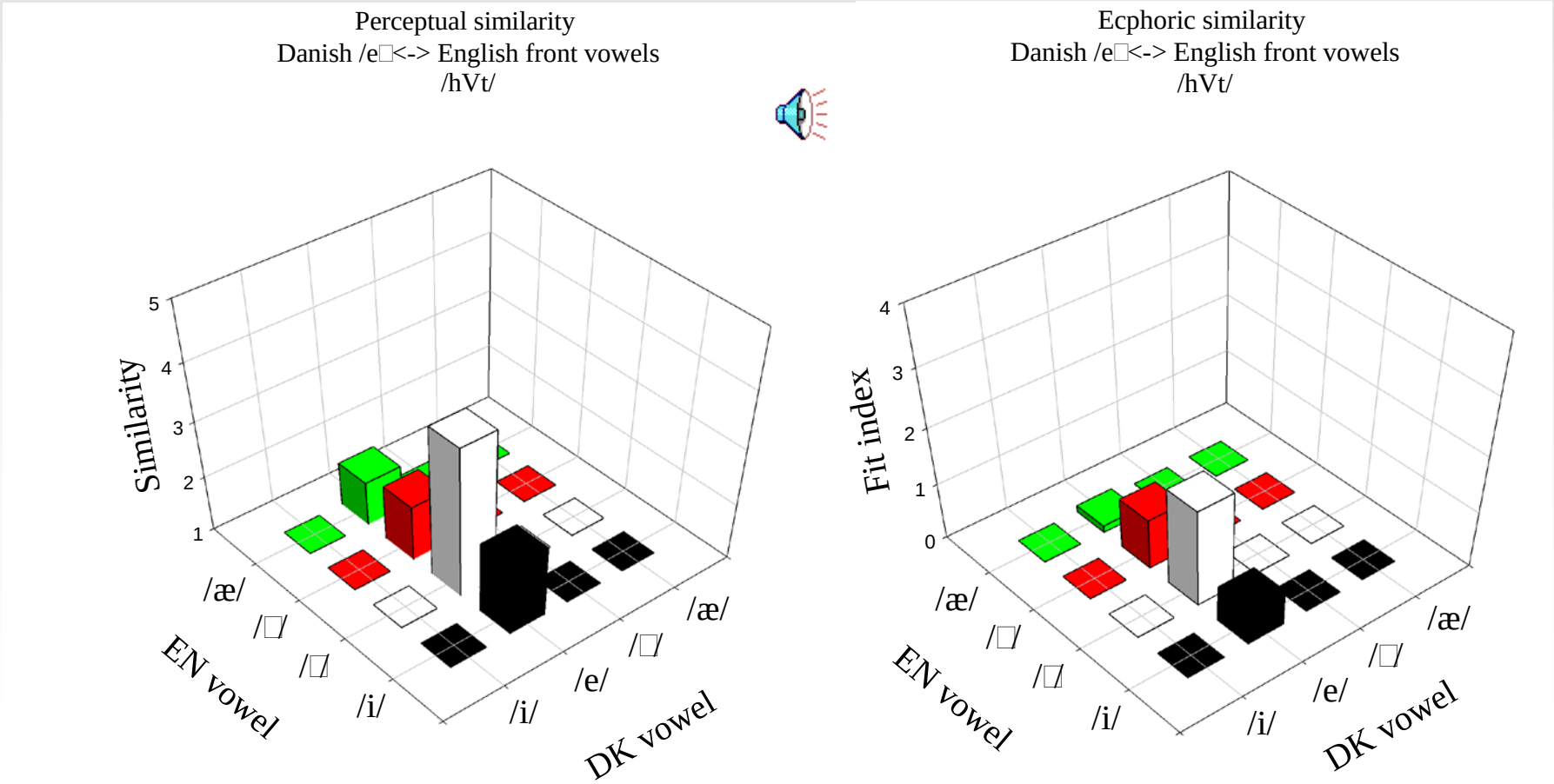
Fit index
p identification x goodness rating



Comparison of perceptual and ecphoric similarity

Graded discrimination

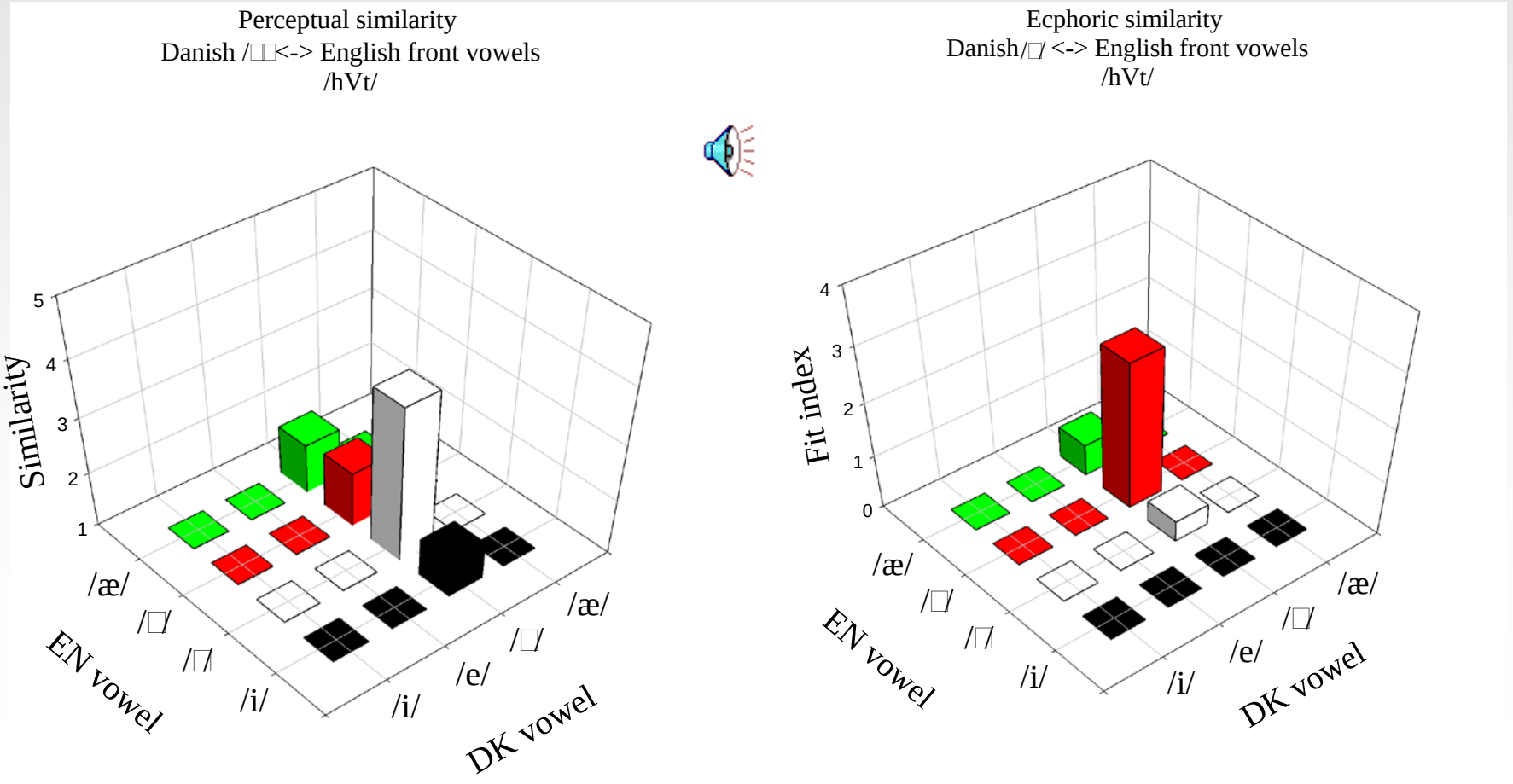
Fit index
p identification x goodness rating



Comparison of perceptual and ecphoric similarity

Graded discrimination

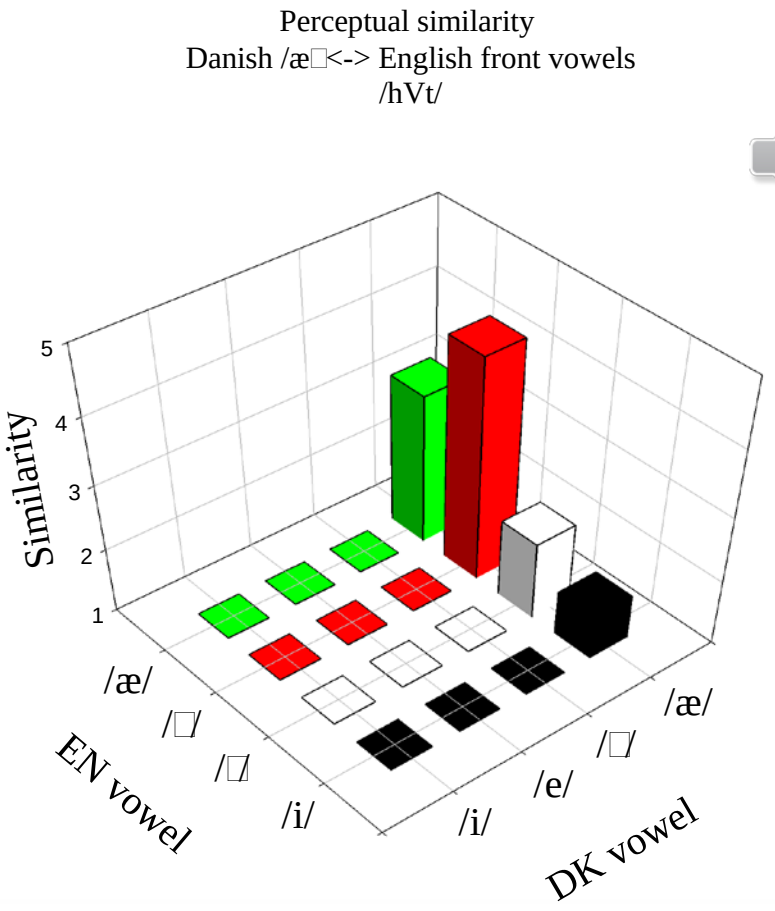
Fit index
p identification x goodness rating

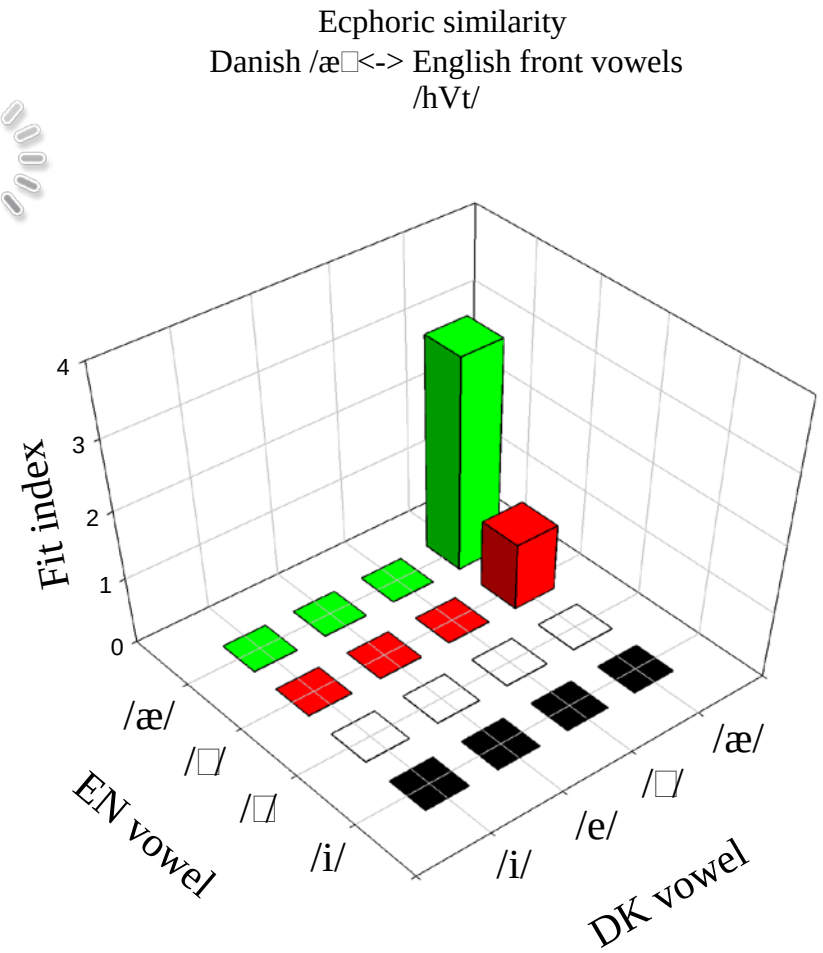


Comparison of perceptual and ecphoric similarity

Graded discrimination

Fit index
p identification x goodness rating



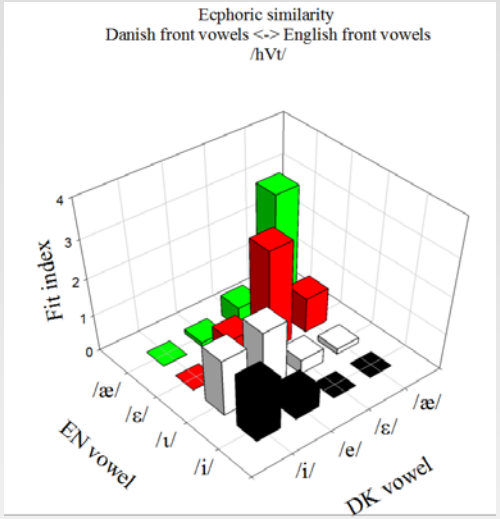


Comparison of ecphoric and perceptual similarity

Task-dependent differences:

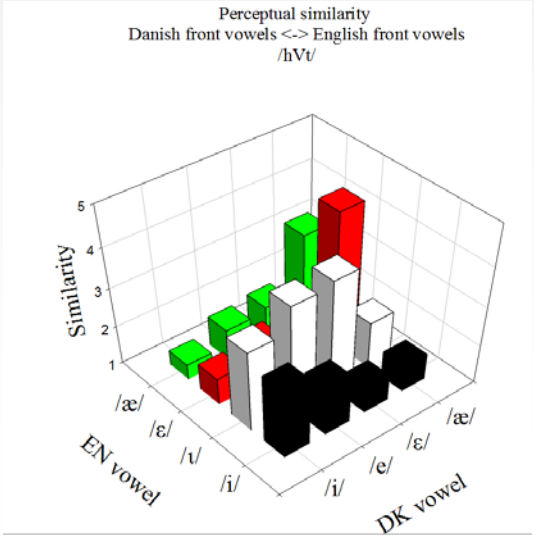
Ecphoric similarity:
Specific L2 token -> L1 category

Clean data: Category match



Perceptual similarity:
Specific L2 token <-> Specific L1 token

Rich (noisy?) data: All L1 – L2 sound pairs



Comparison of ecphoric and perceptual similarity

Ecphoric similarity: Specific L2 token -> L1 category

Problem:
Orthographic (etc.) response = internal representation of L1 category?

Perceptual similarity: Specific L2 token <-> Specific L1 token

Problem:
Specific L1 production = listerners' internal representation?

If orthographic response ≠ listeners' internal representations:

If L1 production ≠ listeners' internal representations:

How well does a) ecphoric b) perceptual similarity predict perceptual difficulty?

How well does a) ecphoric b) perceptual similarity predict perceptual difficulty?

AXB discrimination experiments DK /i-e/, /i-ɛ/, /e-ɛ/ (36 trials each)

Subjects:10 L1 SBE listeners (2f, 8m, M_{age} = 29.1), minimal L2 Danish experience)

Predictions:

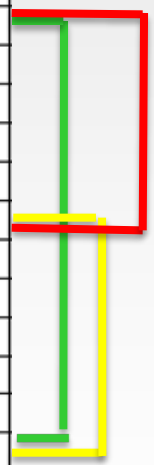
ecphoric similarity:

How well does a) ecphoric b) perceptual similarity predict perceptual difficulty?

AXB discrimination experiments DK /i-e/, /i-ɛ/, /e-ɛ/ (36 trials each)

Subjects:10 L1 SBE listeners (2f, 8m, M_{age} = 29.1), minimal L2 Danish experience)

DK vowel	SBE category/vowel	Fitindex
/i/	/i/	1.6
	/ɪ/	**1.7**
	/ɛ/	0
	/æ/	0
/e/	/i/	0.2
	/ɪ/	**2.6**
	/ɛ/	1.05
	/æ/	0.4
/ɛ/	/i/	0
	/ɪ/	0.6
	/ɛ/	**2.7**
	/æ/	1.1



Predictions:

ecphoric similarity:

TC CG

/i-ɛ/ = /e-ɛ/ < /i-e/

easy _____ hard

How well does a) ecphoric b) perceptual similarity predict perceptual difficulty?

AXB discrimination experiments DK /i-e/, /i-ɛ/, /e-ɛ/ (36 trials each)

Subjects:10 L1 SBE listeners (2f, 8m, M_{age} = 29.1), minimal L2 Danish experience)

DK vowel	SBE category/vowel	Fitindex	Similarity rating
/i/	/i/	1.6	2.8
	/ɪ/	**1.7**	**3.1**
	/ɛ/	0	1.7
	/æ/	0	1.4
/e/	/i/	0.2	2.3
	/ɪ/	**2.6**	**3.7**
	/ɛ/	1.05	2
	/æ/	0.4	1.8
/ɛ/	/i/	0	1.7
	/ɪ/	0.6	**3.8**
	/ɛ/	**2.7**	2
	/æ/	1.1	1.9

Predictions:

ecphoric similarity:

TC CG

/i-ɛ/ = /e-ɛ/ < /i-e/

easy _____ hard

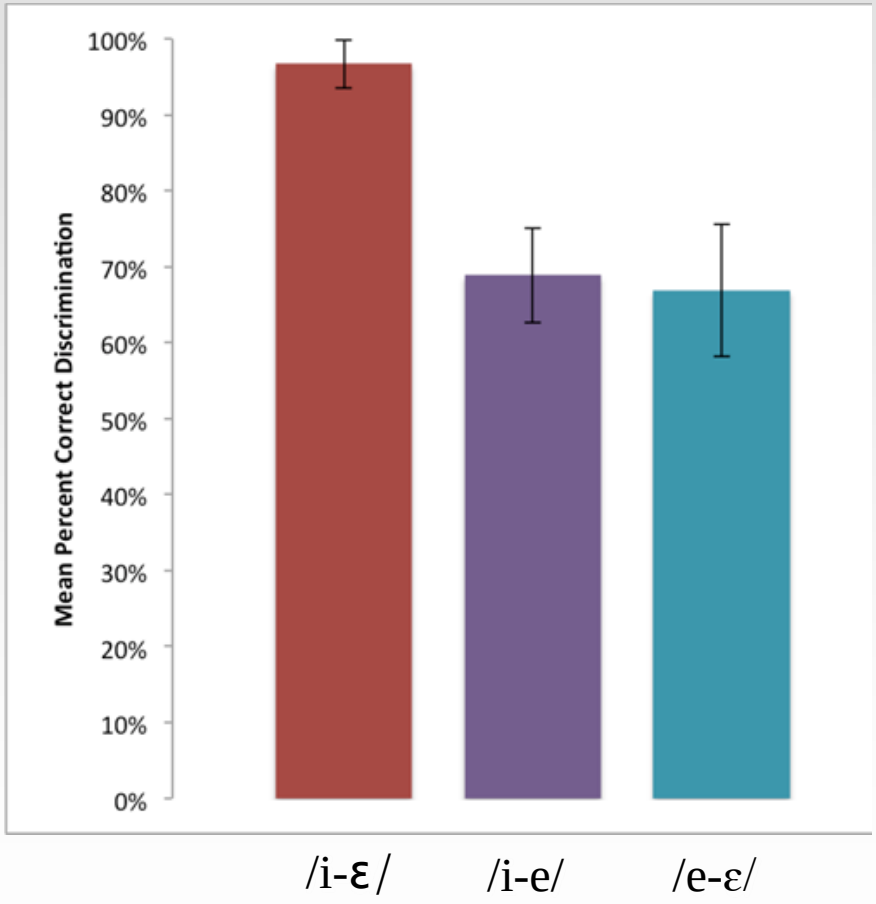
perceptual similarity:

CG SC

/i-ɛ/ ≤ /i-e/ < /e-ɛ/

easy _____ hard

How well does a) ecphoric b) perceptual similarity predict perceptual difficulty?



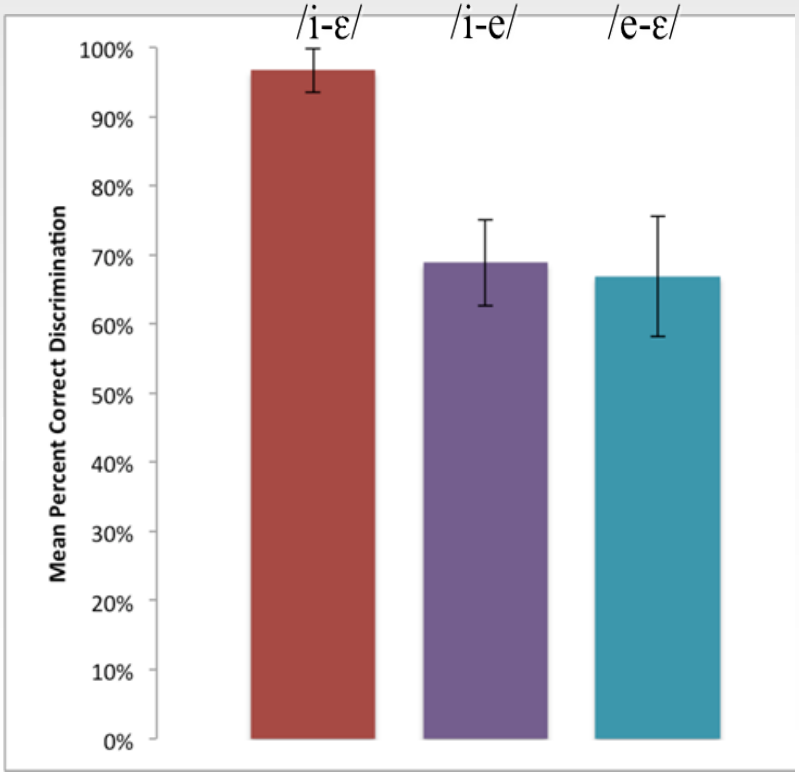
Result of AXB discrimination

/i-ɛ/ < /i-e/ = /e-ɛ/

How well does a) ecphoric b) perceptual similarity predict perceptual difficulty?

Somewhat frustrating response: /i-ɛ/ < /i-e/ = /e-ɛ/

Both types of similarity yield correct & incorrect predictions



ecphoric similarity:

TC CG

/i-ɛ/ = /e-ɛ/ < /i-e/

easy _____ hard

perceptual similarity:

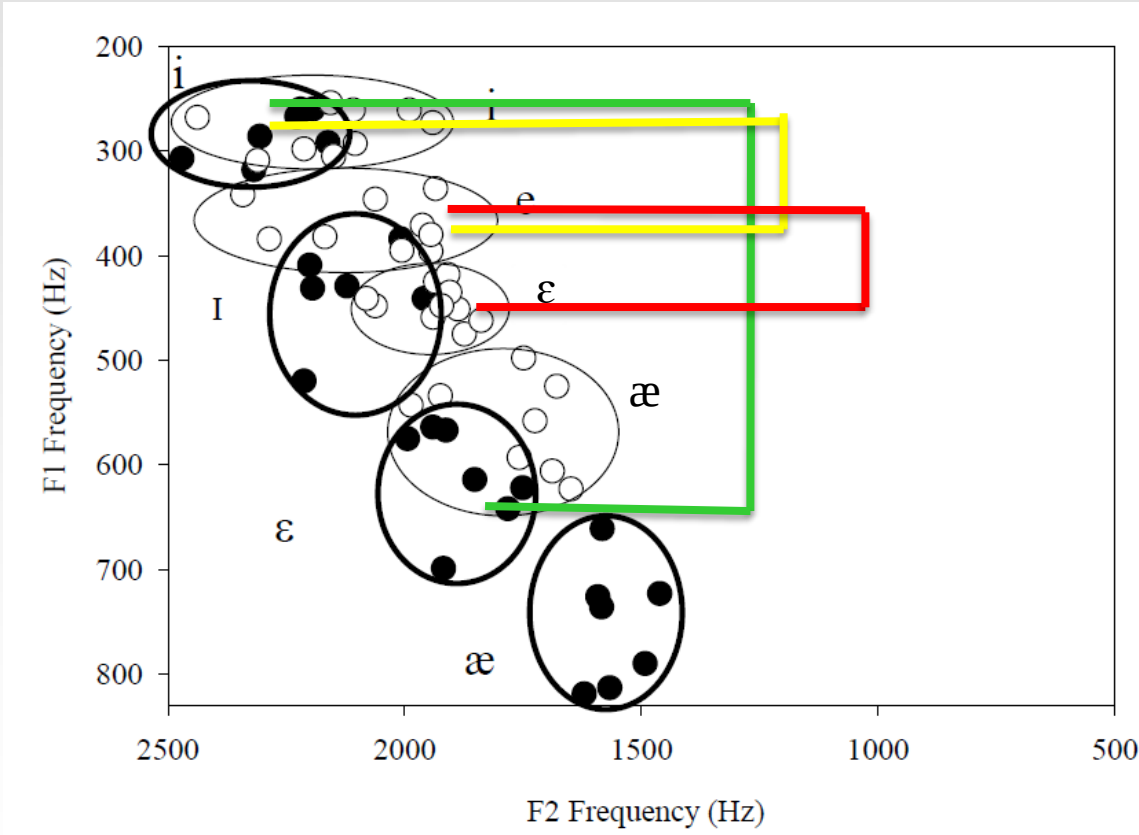
CG SC

/i-ɛ/ ≤ /i-e/ < /e-ɛ/

easy _____ hard

Any help from acoustics?

F1/F2 frequencies English and Danish vowels used in study



Predictions from acoustic comparison:

/i-ɛ/ < /i-e/ < /e-ɛ/

Any help from acoustics?

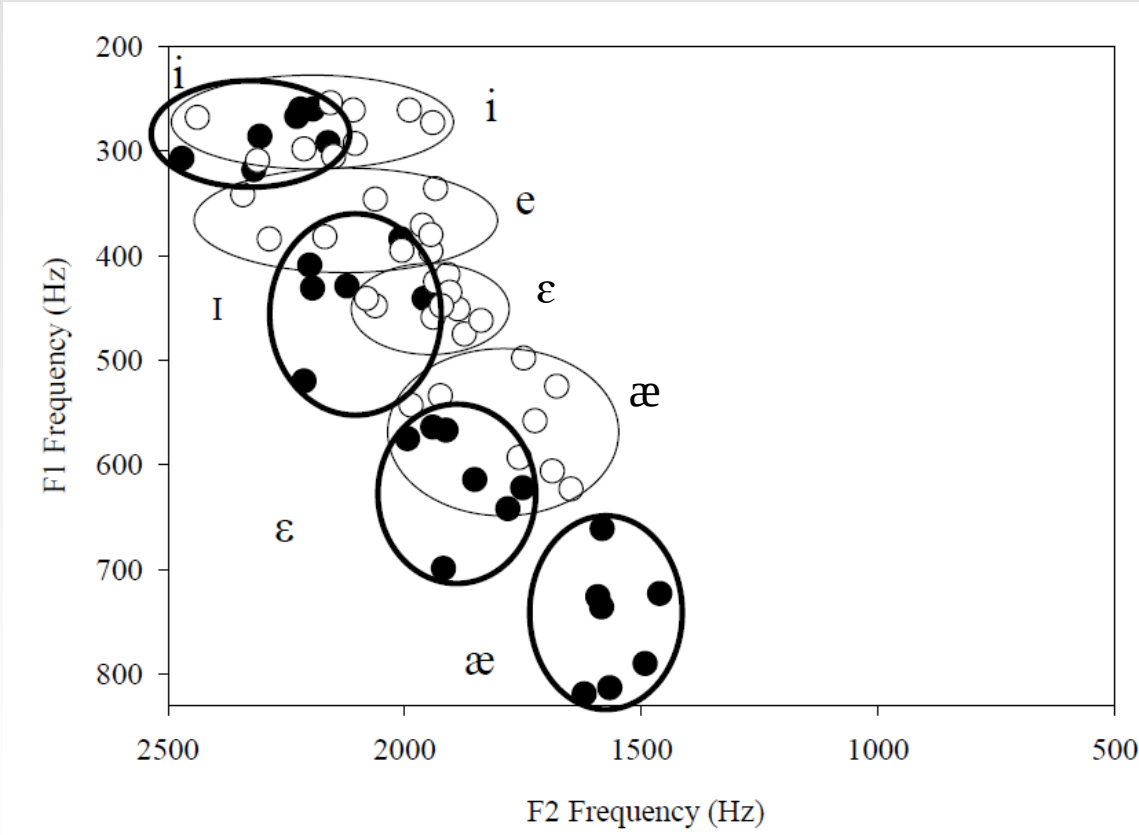
Perceptual (discrimination) difficulty:

/i-ɛ/ < /i-e/ = /e-ɛ/

Predictions from acoustic comparison:

/i-ɛ/ < /i-e/ < /e-ɛ/

F1/F2 frequencies English and Danish vowels used in study



Conclusion
Predicting difficulty in non-native speech perception and production

How:

- Comparison of symbols (”armchair method”): At best a very first heuristic of limited use, may be misleading
- Acoustic comparisons:
  - Can yield useful predictions (if lucky)
  - Likely to run into problems (normalization, phonetic, phonotactic, prosodic context, different effects of these contexts across languages and relation to perception, lack of compatibility with perception)

Conclusion
Predicting difficulty in non-native speech perception and production

How:

- Two ways of assessing perceived similarity
  - Perceptual similarity: Comparison of stimulus dyads (”overt”)
  - Ecphoric similarity: Comparison of stimulus to internal representation

No clear winner

”But of course similarity is a vague concept.” (Greenberg 1966)